VOLUME 107 PART B NUMBER 34 JULY 1960

The Proceedings

OF

THE INSTITUTION OF ELECTRICAL ENGINEERS

FOUNDED 1871: INCORPORATED BY ROYAL CHARTER 1921

PART B

ELECTRONIC AND COMMUNICATION ENGINEERING
(INCLUDING RADIO ENGINEERING)

SAVOY PLACE · LONDON W.C.2

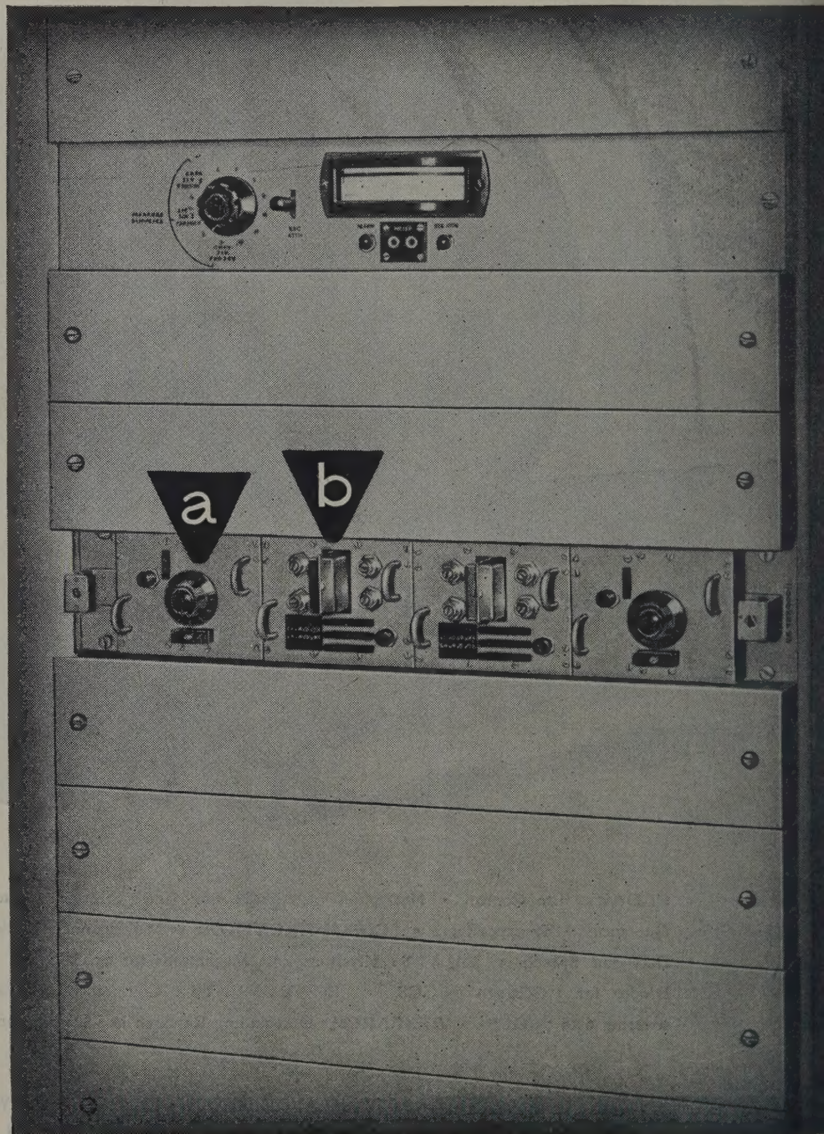
Price Ten Shillings and Sixpence

NEW TRANSISTORIZED CARRIER

For Cable Systems · Radio Systems ·

TECHNICAL FEATURES

- 300-3,400 c/s Channel Bandwidth.
- 4 kc/s spaced.
- Optional Inbuilt Out-band signalling (3825 c/s at -20 dbmO)
- Fully transistorized.
- 96 Circuits without signalling or 48 circuits with signalling per 9 ft. rackside (20½ in. x 8½ in. floor dimensions).
- Compact plug-in units with hinged, card-mounted components give maximum component accessibility.
- Can be supplied on racksides or a complete basic group on a sub-frame.
- Available in 60-108 kc/s basic channel groups or as systems for 12-circuit 2-wire cable operation or multi-circuit, for radio links.
- Conforms with CCITT and BPO requirements.



Cover removed to show panel with Channel and Signalling Units in posit

TMC

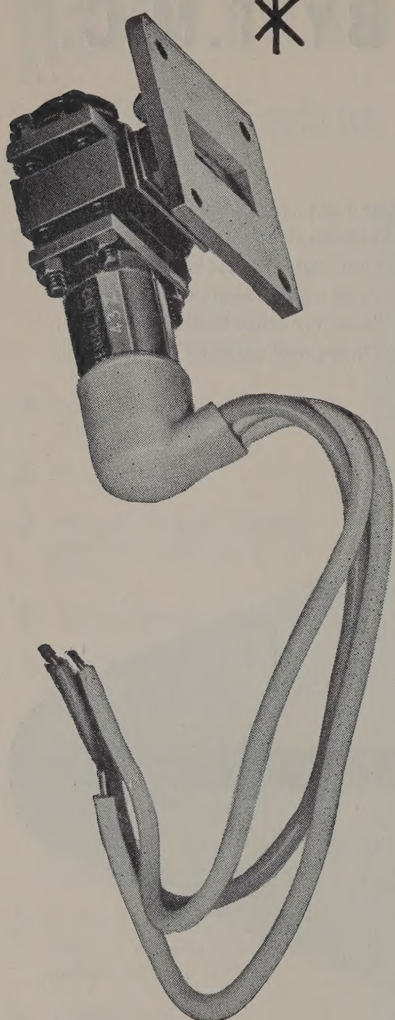
Further information from:

TELEPHONE MANUFACTURING COMPANY LIMITE

Transmission Division: Cray Works, Sevenoaks Way, Orpington, Ken

Telephone: Orpington 26611

SELLING AGENTS *Australia and New Zealand:* Telephone Manufacturing Co. (A'sia) Pty Ltd., Sydney, New South Wales.
Canada and U.S.A.: Telephone Manufacturing Co. Ltd., Toronto, Ontario.
All other Countries (for transmission equipment only): Automatic Telephone and Electric Co. Ltd., Lond



QUIET AS A MOUSE! OUR FORCED AIR-COOLED **K 350** A TWO RESONATOR OSCILLATOR KLYSTRON

Specially designed for low noise characteristics, the K350 is just one of the large range of Klystrons manufactured by the **English Electric Valve Co. Ltd.** It operates in the 8500 - 10000 Mc/s range and has mechanical tuning. For data and information concerning any of our wide range of thermionic tubes for industry, write to the Company

'ENGLISH ELECTRIC'

AGENTS THROUGHOUT THE WORLD

ENGLISH ELECTRIC VALVE CO. LTD.

Chelmsford, England
Telephone: Chelmsford 3491


QUEBEC
KEY
TERMINAL
REPEATER 2000 Mc/s RADIO CHANNEL
FOR 300 CHANNELS
OR COLOUR TELEVISION
MULTIPLEX EQUIPMENT
MPBELLTON
BATHURST
NEWCASTLE
EW
RUNSWICK
PRINCE
EDWARD
ISLAND
SYDNEY
SCOTCH
MOUNTAIN
REPEATER
MONCTON
SAINT
JOHN
49 MILE
PATH
OVER WATER
HALIFAX
NOVA
SCOTIA

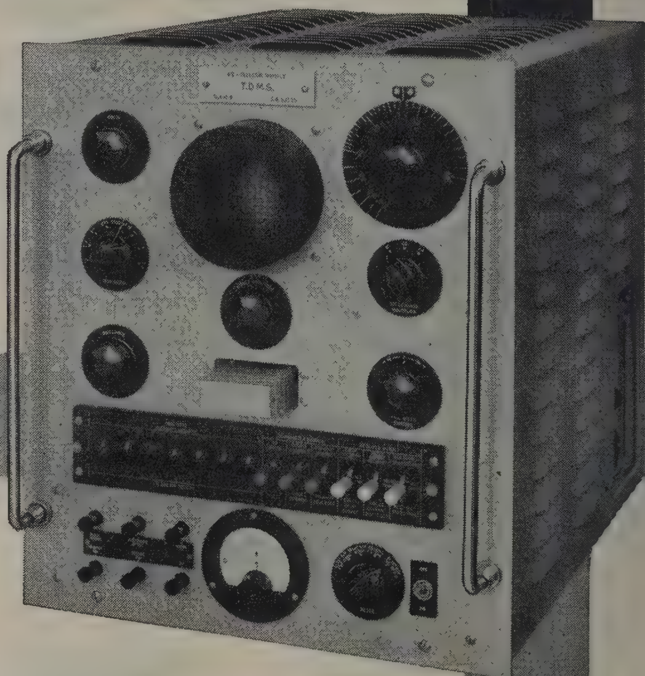
1:1
2:2
6:1
7:7
Q9S
or any single character

Telegraph distortion measurement

in the speed range 40-400 bauds

Rapidly increasing international use of high-speed synchronous telegraph systems has led to the urgent need for suitable distortion measuring equipment. M.S. Type TCA1219, manufactured by A.E.I. at their Woolwich factory to a British Post Office design, enables distortion measurements to be made within 1% accuracy. The method of indication is by bright spots on the circular time base of a cathode-ray tube. Screening and the fitting of radio interference suppressors have been given particular attention.

Size	17 in. high x 15 in. wide x 18 in. deep
Weight	110 lb.
Power supply	200/250 v. 50 c.p.s.
Power consumption	170 w.



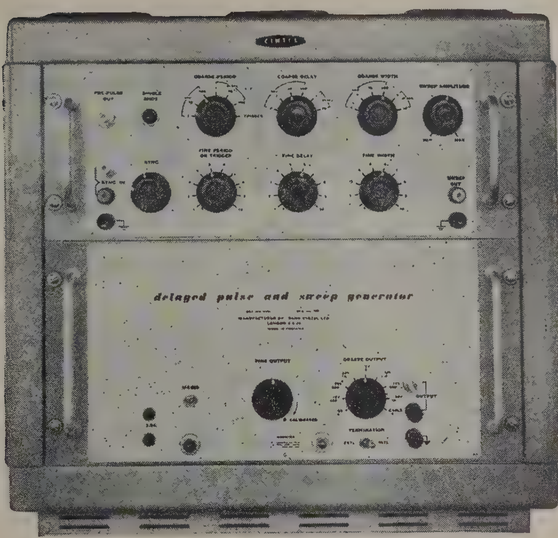
vast experience in the science of telecommunications

AEI **Telecommunications Division**
Transmission Department Woolwich London SE18
Associated Electrical Industries Limited

TA 3795


Whole
in one
PART NO.
26/A/14
It's a comforting thought that a complete wiring system, embracing perhaps five hundred connections, can be bought-in simply by quoting one part number.
Preformed wiring by Plessey, in conjunction with Plessey plugs and sockets, solves many of the problems encountered in the making up of wiring assemblies used in the manufacture of electronic and electrical equipment, and the benefits both structurally and economically are considerable.
One order on Plessey for a complete assembly—or set of assemblies—means only one supplier to progress—only one item to be dealt with by the various departments concerned.
Let us begin at the beginning
In providing complete wiring forms, cable assemblies and junction boxes to special requirements, the maximum economy can be achieved by calling in Plessey at the design stage. In this way can the wide experience of the engineers be used to your full advantage.
Plessey
WIRING & CONNECTORS DIVISION
THE PLESSEY COMPANY LIMITED · CHENEY MANOR · SWINDON · WILTS
Telephone: Swindon 6251
Overseas Sales Organisation: Plessey International Limited · Ilford · Essex · Tel: Ilford 3040
CW 11A

DELAYED PULSE AND SWEEP GENERATOR



A versatile pulse generator designed to meet the need for a comprehensive instrument covering a wide range of pulse work. Four main facilities are provided: a pre-pulse, a main pulse delayed on the pre-pulse, a negative going sawtooth and a fast rising pulse formed from a pure line.

BRIEF SPECIFICATION

Period

Continuously variable from 0·9μsec to 1·05sec i.e. 0·95c/s to 1·1Mc/s. Accuracy ±5%.

Pre-pulse

40mμsec. 8V peak in 75Ω, positive going.

Main pulse

Width: Variable from 0·09μsec to 105msec ±5%.

Amplitude: Control gives 4:1 attenuation of each of four maximum outputs as follows:

5V max in 75Ω	rise time	10mμsec
10V max in 150Ω	rise time	<20mμsec
25V max in 600Ω	rise time	<40mμsec
50V max in 1000Ω	rise time	50mμsec

Polarity: Positive or negative going.

Accuracy: ±2%.

Delay

Conclusion of pre-pulse to advent of main pulse, delay variable from 0·09μsec to 105msec. Accuracy ±5%.

Sweep

D.C. coupled negative going sawtooth same width and delay as main pulse.
15V peak max.

Cable pulse

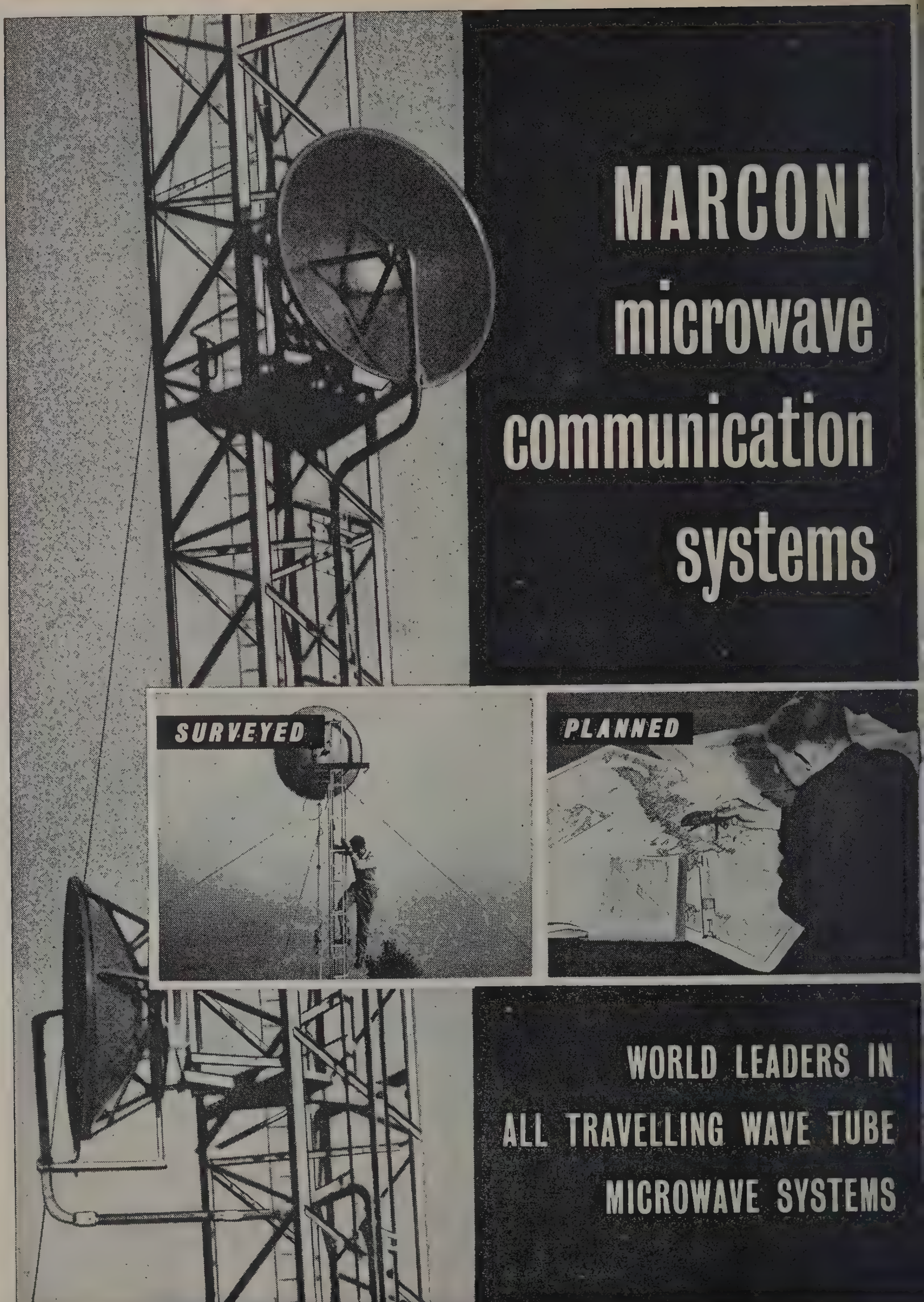
Obtained from short circuited pure line.
One positive and one negative going pulse coincident with main pulse.
25mμsec wide 3V max in 75Ω, rise time <8mμsec.

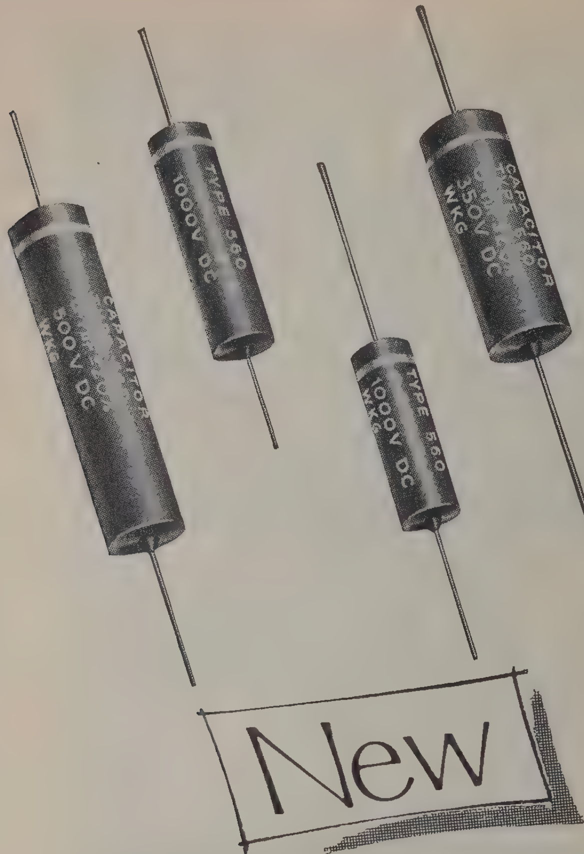
Sync, trigger or single shot facilities provided.
Full data available on request.

CINTEL INSTRUMENT Division

RANK CINTEL LIMITED
WORSLEY BRIDGE ROAD · LONDON · SE26
HITHER GREEN 4600

Sales and Servicing Agents: Atkins, Robertson & Whiteford Ltd., Industrial Estate, Thornliebank, Glasgow.
McKellen Automation Ltd., 122 Seymour Grove, Old Trafford Manchester, 16. Hawnt & Co., Ltd., 59 Moor St., Birmingham, 4.


MARCONI
microwave
communication
systems
SURVEYED
PLANNED
WORLD LEADERS IN
ALL TRAVELLING WAVE TUBE
MICROWAVE SYSTEMS



- **No exposed metal parts other than terminations, which are clean solder coated, thereby ensuring easy soldering.**
- **Body and terminations free of wax coating or any other low melting point material.**
- **Long life without voltage derating.**
- **Designed to meet the requirements of British Joint Service Standards RCS 131 and BS 2131 with humidity classification H.2.**
- **Solid construction eliminates internal movement, preventing damage by severe vibration.**

DUBILIER ENCAPSULATED PAPER DIELECTRIC TUBULAR CAPACITORS HAVING OUTSTANDING CHARACTERISTICS

The Dubilier Capacitor Type 560 is a new approach to capacitor requirements for all radio and electronic applications. It is constructed to meet long and arduous service conditions. The paper dielectric element is impregnated with a plastics material to produce a solid unit. The terminations are of great mechanical and electrical strength and the assembled element is sealed in an encapsulated mineral loaded epoxy resin so that there are no parts capable of movement, making the capacitor completely immune to shock and all normal atmospheric conditions.
Capacitance Tolerance; ±20% normal ±10% by selection. Power Factor; Less than 1% at 1,500 c/s. Insulation Resistance; Better than 20,000MΩ at normal temperature. Voltage Application; From—40° to +125°C for d.c. and from —40° to +70°C for a.c.

CAPACITANCE μF	VOLTAGE RATINGS			DIMENSIONS	
	d.c. Wkg. at −40°C to +125°C	d.c. Test at 20°C	a.c. Wkg. r.m.s. at −40°C to +70°C and up to 60 c/s	Diameter +0.020″ −0	Length ±0.040″
0.001	1,000	2,500	250	3/8	1
0.002	1,000	2,500	250	3/8	1
0.005	1,000	2,500	250	3/8	1
0.01	1,000	2,500	250	3/8	1 3/8
0.02	750	2,250	250	3/8	1 3/8
0.05	500	1,500	250	1/2	1 3/8
0.1	350	1,000	180	1/2	1 3/8
0.1	500	1,500	250	1/2	1 15/16

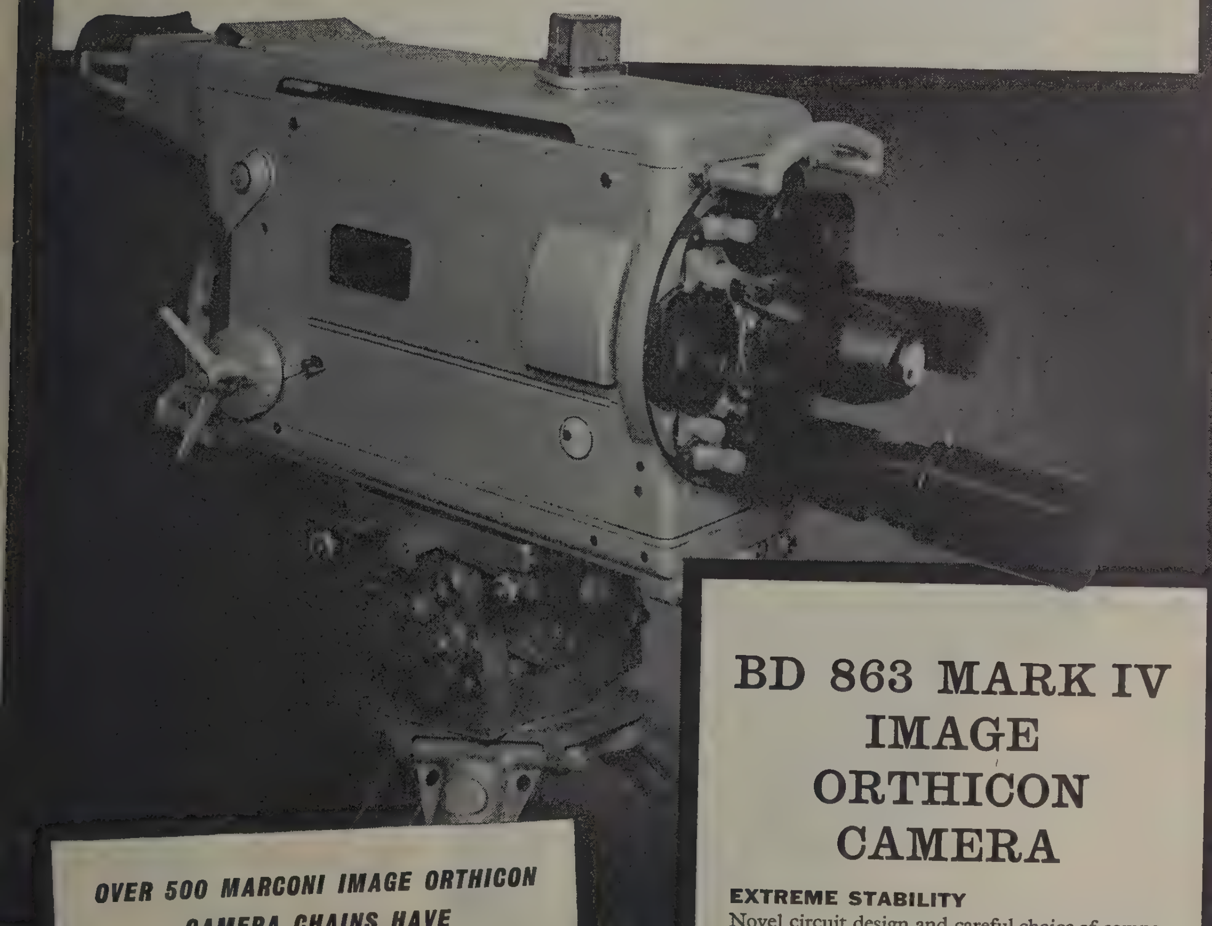
DUBILIER

DUBILIER CONDENSER CO. (1925) LIMITED, DUCON WORKS, VICTORIA ROAD, NORTH ACTON, LONDON W.3
Telephone: ACOrn 2241 (5 lines)
Cables: Hivoltcon London
Telegrams: Hivoltcon London Telex
Telex: 25373
DN 242A

THE MARK IV CAMERA CHAIN

EXPERIENCE COUNTS

Marconi's pioneered the use of the 4½ inch Image Orthicon Camera using the tube developed by their associates, the English Electric Valve Company. Marconi's have amassed more 'know-how' on the use of the 4½ inch Image Orthicon than any other manufacturer.



OVER 500 MARCONI IMAGE ORTHICON CAMERA CHAINS HAVE BEEN SOLD THROUGHOUT THE WORLD

BD 863 MARK IV IMAGE ORTHICON CAMERA

EXTREME STABILITY
Novel circuit design and careful choice of components gives such a high degree of stability that operational controls have been removed from the camera.

FIRST CLASS PICTURE QUALITY
The 4½ inch Image Orthicon tube gives a picture quality substantially better than any other type or size.

LIGHT AND COMPACT
By reducing and simplifying the camera electronics its weight has been held below 100 lb. and its size made correspondingly small.

MARCONI

COMPLETE SOUND AND TELEVISION SYSTEMS

MARCONI'S WIRELESS TELEGRAPH COMPANY LIMITED · CHELMSFORD · ESSEX · ENGLAND

M4D

The Ferranti 3 DIGIT VOLTMETER Type D101

The need for instruments capable of measuring voltages with a high degree of accuracy and with a fast reading time has long been apparent. The Ferranti 3 digit voltmeter has been developed to meet this requirement. The advantages of this precision instrument will undoubtedly prove attractive to those engaged in the fields of automatic testing and monitoring, analogue to digital conversion, calibration of moving pointer instruments and many similar applications.

Special Features

- **Automatic Ranging and Polarity**
- **High Accuracy and Resolution**
- **Fast Reading Time**
- **Complete Reliability**

SPECIFICATION

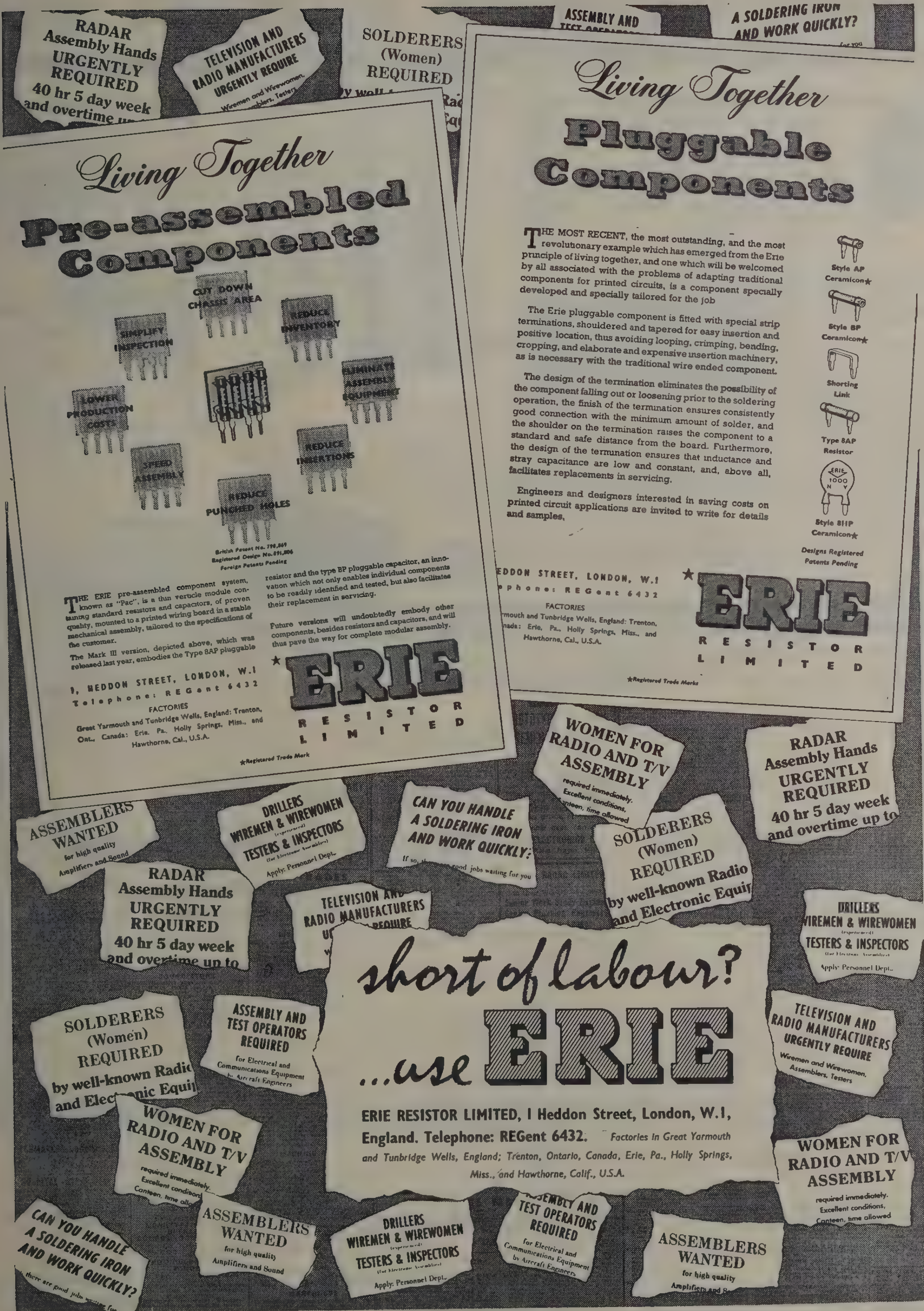
Display	Three digit plus automatic polarity indication and automatic decimal placement.
Automatic Ranges	D.C. Volts in 3 ranges 0.01 — 9.99 V 10.0 — 99.9 V 100V — 999 V
Accuracy	0.1% of full scale reading on any range.
Average Reading Time	0.7 seconds.
Input Impedance at Balance	10 Megohms.
Input	110 — 250V A.C.50—60 c/s 50W.
Weight	50 lbs. approximately.
Style	Bench cabinet 17″ x 13″ x 10½″ high with optional brackets for standard rack mounting.

In view of continuing development, the right is reserved to alter the specification or design of this instrument.

FERRANTI LTD · FERRY ROAD · EDINBURGH 5

Telephone: DEAn 1211

ES/T64


Living Together
Pre-assembled Components
CUT DOWN CHASSIS AREA
SIMPLIFY INSPECTION
REDUCE INVENTORY
ELIMINATE ASSEMBLY EQUIPMENT
LOWER PRODUCTION COSTS
REDUCE INSERTIONS
SPEED ASSEMBLY
REDUCE PUNCHED HOLES
British Patent No. 798,869
Registered Design No. 891,806
Foreign Patents Pending
THE ERIE pre-assembled component system, known as "Pac", is a thin verticle module containing standard resistors and capacitors, of proven quality, mounted to a printed wiring board in a stable mechanical assembly, tailored to the specifications of the customer.
The Mark III version, depicted above, which was released last year, embodies the Type 8AP pluggable resistor and the type BP pluggable capacitor, an innovation which not only enables individual components to be readily identified and tested, but also facilitates their replacement in servicing.
Future versions will undoubtedly embody other components, besides resistors and capacitors, and will thus pave the way for complete modular assembly.
1, HEDDON STREET, LONDON, W.1
Telephone: REGent 6432
FACTORIES
Great Yarmouth and Tunbridge Wells, England: Trenton, Ont., Canada: Erie, Pa., Holly Springs, Miss., and Hawthorne, Cal., U.S.A.
★ ERIE RESISTOR LIMITED
★Registered Trade Mark
Living Together
Pluggable Components
THE MOST RECENT, the most outstanding, and the most revolutionary example which has emerged from the Erie principle of living together, and one which will be welcomed by all associated with the problems of adapting traditional components for printed circuits, is a component specially developed and specially tailored for the job
The Erie pluggable component is fitted with special strip terminations, shouldered and tapered for easy insertion and positive location, thus avoiding looping, crimping, bending, cropping, and elaborate and expensive insertion machinery, as is necessary with the traditional wire ended component.
The design of the termination eliminates the possibility of the component falling out or loosening prior to the soldering operation, the finish of the termination ensures consistently good connection with the minimum amount of solder, and the shoulder on the termination raises the component to a standard and safe distance from the board. Furthermore, the design of the termination ensures that inductance and stray capacitance are low and constant, and, above all, facilitates replacements in servicing.
Engineers and designers interested in saving costs on printed circuit applications are invited to write for details and samples.
Style AP Ceramicon★
Style BP Ceramicon★
Shorting Link
Type 8AP Resistor
ERIE 1000 N V
Style 811P Ceramicon★
Designs Registered
Patents Pending
EDDON STREET, LONDON, W.1
phone: REGent 6432
FACTORIES
rmouth and Tunbridge Wells, England: Trenton,
nada: Erie, Pa., Holly Springs, Miss., and Hawthorne, Cal., U.S.A.
★ ERIE RESISTOR LIMITED
★Registered Trade Marks
RADAR Assembly Hands URGENTLY REQUIRED 40 hr 5 day week and overtime up to
TELEVISION AND RADIO MANUFACTURERS URGENTLY REQUIRE Wiremen and Wirewomen, Assemblers, Testers
SOLDERERS (Women) REQUIRED by well-known Radio and Electronic Equip
WOMEN FOR RADIO AND T/V ASSEMBLY required immediately. Excellent conditions, Canteen, time allowed
ASSEMBLERS WANTED for high quality Amplifiers and Sound
DRILLERS WIREMEN & WIREWOMEN (experienced) TESTERS & INSPECTORS Apply: Personnel Dept.
CAN YOU HANDLE A SOLDERING IRON AND WORK QUICKLY?
ASSEMBLY AND TEST OPERATORS REQUIRED for Electrical and Communications Equipment by Aircraft Engineers
short of labour? ...use ERIE
ERIE RESISTOR LIMITED, 1 Heddon Street, London, W.1, England. Telephone: REGent 6432. Factories in Great Yarmouth and Tunbridge Wells, England; Trenton, Ontario, Canada, Erie, Pa., Holly Springs, Miss., and Hawthorne, Calif., U.S.A.

INDEX OF ADVERTISERS

THE PROCEEDINGS OF
THE INSTITUTION OF ELECTRICAL ENGINEERS

EDITED UNDER THE SUPERINTENDENCE OF W. K. BRASHER, C.B.E., M.A., M.I.E.E., SECRETARY

/OL. 107. PART B. NO. 34. JULY 1960

21.317.724: 621.374.32

The Institution of Electrical Engineers
Paper No. 3238 E
July 1960

THE DESIGN AND USE OF INSTRUMENTS FOR COUNTING LOCAL LIGHTNING FLASHES

By F. HORNER, M.Sc., Member.

(*The paper was first received* 22*nd May*, 1959, *and in revised form* 26*th January*, 1960.)

SUMMARY

Instruments for counting local lightning flashes are reviewed and easons are given for selecting one particular design for extensive ests. The importance of the aerial configuration is stressed. A 7 m ertical aerial has been adopted as a standard, and the instruments ave normally been adjusted to trigger on a 3-volt signal from a built-in alibrator, corresponding to a field change of about 3 volts/m.

With this arrangement, on a day on which thunder is heard, there s usually a maximum hourly count exceeding 30 and a total daily ount exceeding 100.

The effective range of a counter is defined in statistical terms and echniques for its measurement are discussed. Preliminary estimates f the effective range are derived from observations on a few local torms in England; with the standard instrument it is about 30 km. More observations are required to confirm the results, particularly in ountries with frequent thunderstorms. Equipment should be standardized so far as possible.

(1) INTRODUCTION

The geographical distribution of the occurrence of thunderstorms is of interest to both electrical power and radio engineers. Power engineers wish to know the probability of strokes to arth in a given area in order to assess the chances of damage o transmission lines. Radio engineers are interested in thunderstorms as generators of radio noise, and since both earth strokes nd cloud strokes are significant, their requirements differ from hose of power engineers. It is mainly the radio application which is considered here.

Existing knowledge of the occurrence of thunderstorms on a world-wide basis is derived from aural observations of thunder t meteorological stations. Brookes,[1] in 1925, carried out the rst general survey of data, and published maps of the distribuion of thunderstorm days, i.e. the number of days on which hunder is heard at each location. Tables of the average umber of thunderstorm days in each month of the year at a arge number of stations have recently been published by the World Meteorological Organization.[2] This information has serious limitations for application to the radio-noise problem, however, and consideration has therefore been given to the possibility of improving data by the use of some automatic means for recording the incidence of local thunderstorms. Some experiments with instruments designed for the purpose are described in the paper.

(2) REQUIREMENTS FOR THUNDERSTORM DATA FOR RADIO PURPOSES

In a recently published report[3] estimates of radio-noise power received from thunderstorms were given for all times and all places in the world. These were based partly on measurements of noise at about 30 stations, but the general form of the distributions was based on thunderstorm observations at a much larger number.[2] It was possible to establish an approximate relationship between thunderstorm days and radio noise, but the thunderstorm day, as an index of activity, suffers from the following deficiencies:

(*a*) It does not show the diurnal variations of thunderstorm activity.
(*b*) It does not give an indication of intensity or duration.
(*c*) The audibility of thunder is dependent on local conditions and on the personal characteristics of the observer.
(*d*) The range of audibility is so small that a thunderstorm day is a rare event over much of the world.

The need for accurate information on both the geographical distribution of thunderstorms and the nature of the electromagnetic radiation from them is likely to increase, particularly for the estimation of radio noise received on a highly-directional aerial which may or may not be directed towards a storm centre.

Proposals have been made from time to time for an automatic recorder of local lightning, and suggested instruments have taken the form of a simple radio receiver, the output from which has actuated a counting device or register. Before comparing the various designs it is necessary to discuss briefly some of the main features of the lightning flashes which it is proposed to count. For a fuller discussion reference may be made to one of several recent surveys, for example that by Schonland.[4]

In most lightning flashes there occurs a quasi-continuous form of discharge lasting up to about half a second, and this prolonged discharge is the main source of radio noise at high frequencies.

Written contributions on papers published without being read at meetings are vited for consideration with a view to publication.
The paper is an official communication from the Radio Research Station, Department of Scientific and Industrial Research.

It may take place entirely within the cloud or may extend downwards towards the earth, establishing an ionized channel. If it reaches the earth a relatively large current—about 10 kA—flows between the ground and the charged cloud. This discharge, known as the return stroke, is the main source of radio noise at very low frequencies. The general view at present is that in a lightning flash which does not reach the ground there is no large-scale current equivalent to the return stroke,[5] although the contrary opinion has been expressed.[6]

If only ground strokes are of interest, as in assessing the likelihood of damage to power lines, a stroke counter need respond only to very low frequencies; some counters have been designed specifically for this purpose. When an instrument is to serve as an indicator of the radio noise generated in lightning it should respond also to higher frequencies. Because there is a significant difference between cloud strokes and ground strokes as generators of noise it would be desirable to count them separately, but if a single counting channel is to be used the data are likely to be more applicable to problems of radio interference if all flashes are counted rather than ground strokes alone.

In some work on the detection of local storms, use has been made of very high frequencies. These have the advantage that propagation is normally restricted to optical paths and in flat country the range of operation tends to be well defined. However, the performance would no doubt depend on the local topography, and since the energy from an atmospheric is small at these frequencies, a counter would be susceptible to man-made interference. All the counters of which details have been published have operated at low or very low frequencies.

Since thunder is heard infrequently in many parts of the world, the range of operation of a counter should preferably exceed that of audible thunder (about 10 km). On the other hand it should not be so large that the records cease to be typical of a particular location. The optimum range is probably about 50 km.

One important factor which enters into discussions of counters is the variation of the field strength of an atmospheric with distance of propagation. At the lowest frequencies and the shortest distances the electrostatic component of field is predominant and an inverse-distance-cubed law applies, but at high frequencies and greater distances where the radiation component predominates the law is that of inverse distance. With a wide-band instrument the law is complex, and one problem is to determine the relationship in various circumstances.

To sum up: the general requirement is for an instrument capable of indicating the total number of lightning flashes within a known radius of about 50 km. Before discussing the means by which this requirement might be achieved, it is useful to review the various designs of counter which have been suggested in the past.

(3) SUMMARY OF EXISTING DESIGNS OF COUNTER

The various counters proposed in recent years are summarized below. Vertical rod aerials were used except where otherwise stated. Since the term 'counter' has been used to denote the complete instrument, the element on which the actual record of the number of counts appears, although often known as a counter, will be called the 'register'.

Forrest (1943).[7]—Forrest recorded the number of atmospherics received on a radio receiver tuned to a frequency in the range 100–150 kc/s, with a bandwidth of 10 kc/s. The aerial length was not specified. A calibration signal was derived from an oscillator. Two counters were used together, with different sensitivities corresponding nominally to ranges of 16 and 160 km. Counts were registered on a recording milliammeter, and data on the diurnal and season variations in the number of coun per hour in England were obtained.

Davis (1946).[8]—In connection with the war-time operation barrage balloons in England, an instrument was designed to gi warning of the approach of storms. It operated on the chan of electrostatic field caused by nearby flashes to ground. A te facility depending on the discharge of a capacitor was include The cable tethering the balloon served as the aerial and the ran varied between 8 km when the balloon was 500 ft high to 25 k when the height was 2000 ft.

Schonland and Gane (1947).[9]—This counter was designed South Africa to record flashes within 11 or 32 km, dependi on the sensitivity setting, and also to give warning of an impe ding near flash by recording unusually large corona current The aerial was 2·4 m long and the response was confined to t lowest frequencies in an effort to record only the major flashe Calibration was effected by applying a d.c. step function.

Gane and Schonland (1948).[10]—The so-called 'ceraunomete was a 2-channel instrument intended to record separately strokes within 10 km and cloud strokes between 10 and 20 kn Very low frequencies were used to obtain an inverse cube la of field strength with distance. Discrimination between t different types of stroke was based on the differing polarity the field changes. A 2·4 m aerial was used and, for calibratio capacitors were discharged through the input circuit.

Foldès (1951).[11]—Foldès compared various instrument including the ceraunometer, and reached the conclusion th in France the best performance was obtained with a narrov band instrument tuned to a frequency near 18 kc/s. In h design a 10 m aerial was used, and the stages were a tune amplifier and a thyratron, operating a register. A charge capacitor was used for calibration. Good correlation w claimed between the counts and meteorological records of loc thunderstorms.

Foldès (1954).[12]—Following the experiments with his 19 design, Foldès found that the instrument could be simplifie without detriment to its performance. In the later design t tuned circuit was connected directly to the thyratron with n intermediate amplifier. Detailed improvements in the metho of calibration were introduced, the basic idea of discharging capacitor through the tuned circuit being retained. Means we provided for recording the rate of counting continuously.

Lugeon (1953).[13]—This was a wide-band v.l.f. instrument wi an amplifying stage, a rectifier and a second biased amplifi which operated a register. Thus the register was operated b an amplified portion of the atmospheric rather than by a trigger electronic circuit. A 50 c/s voltage was used for calibratio With a 20 m aerial satisfactory operation has been claimed c flashes within a radius of 3 km.

Sullivan, Wells and Dinger (1954).[14]—The first count developed at the University of Florida consisted of a filter, amplifying stage, a full-wave rectifier, a second biased amplifi and a register. The nominal bandwidth was from 10 to 28 kc/ A d.c. step-function voltage was used for calibration. With 30 m horizontal aerial 10 m above ground the estimated ran was about 30 km.

Sullivan and Wells (1957).[15]—In a subsequent University Florida design a multivibrator was introduced between t rectifier and the register and the bandwidth was changed to nominal 1–20 kc/s. A 5·5 m vertical aerial was used, and calibr tion was effected by the discharge of a capacitor. Data we obtained on storms up to 150 km range and the requiremen for a range of 30 km were studied.

Ito, Kato and Iwai (1955).[16]—Following tests on a number wide-band and narrow-band counters, these authors recon mended a wide-band v.l.f. instrument consisting of a detecto

amplifier, thyratron and register. Details of the circuit and calibration techniques were not published, but a range of 20 km was said to be obtained. Later (unpublished) work has been carried out in Japan with hard-valve triggered circuits, and a bandwidth of 2–80 kc/s has been recommended.

Pierce (1956).[17]—From a statistical analysis of the electrostatic field changes due to flashes, Pierce discussed the desirable characteristics of counters and suggested a circuit for counting ground strokes within a range of 40 km. It consisted of a simple type of filter, a thyratron or cold-cathode valve and a register. The response was wide-band, the lower limit being about 100 c/s and the upper limit indefinite. The aerial was a long horizontal wire 5 m above ground. Considerable interest in this design has been shown by power engineers.

(4) EXPERIMENTAL COMPARISONS BETWEEN COUNTERS

In 1953 the International Radio Consultative Committee (C.C.I.R.) recommended that, as part of an investigation of radio noise, the available designs of counter should be compared and their relative merits determined. Experiments were started at the Radio Research Station, Slough, and in other countries in 1955. The counters compared at Slough were those of Foldès (1954), Lugeon, and Sullivan and Wells. The design of Gane and Schonland (1948) had previously been used by Foldès and Pierce, who had reported independently that the desired performance was not obtained with European storms. The Forrest instrument operated on a frequency which appeared to be too high to obtain a satisfactory delimitation of the range.

After a series of tests during the summer of 1955 it was concluded that the design of Sullivan and Wells was the most consistent in operation and had a performance most closely approximating to that required. In the Foldès design differences in performance were obtained when different thyratrons of the same type were used. The performance of the Lugeon instrument was satisfactory except that it sometimes failed to register atmospherics of short duration, owing to the slow action of the relay, there being no intermediate triggered-valve circuit. Also calibration with a 50 c/s signal appeared to be less consistent than with a direct step voltage.

The design of Sullivan and Wells, with some modifications, was therefore selected for further tests.

(5) RECOMMENDED DESIGN

Although the Sullivan and Wells design of counter was generally satisfactory, some modifications were made to change the frequency response and to improve stability and convenience of use. The more important features of the design finally adopted are discussed below, and the circuit is shown in Fig. 1. A brief description has already been published.[18]

(5.1) Frequency and Bandwidth

In selecting the frequency response a compromise must be made. The use of the lowest frequencies improves the chances of obtaining a rapid change of field strength with distance and so defining the effective range accurately, but in the present state of knowledge it seems necessary to receive some energy at higher frequencies if cloud flashes are to be recorded. For these reasons a wide-band instrument is preferred, and the nominal response extends from 1 to 20 kc/s. The use of a wide bandwidth also minimizes the effects of differences in the frequency spectrum between one flash and another.

The filter circuit in the recommended design differs somewhat from that proposed by Sullivan and Wells, which was found to have poor response in the lower part of the frequency range when used with a 7 m aerial. The frequency response of the

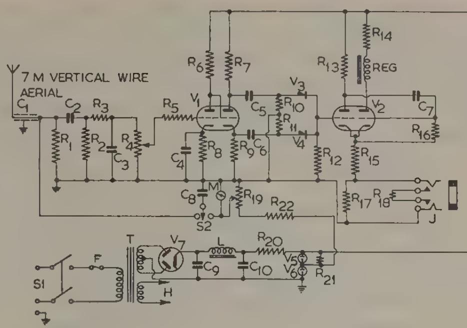


Fig. 1.—Lightning-flash counter.

R_1, R_2, 1 MΩ	R_{13}, 47 kΩ	C_1, 140 pF
R_3, 100 kΩ	R_{14}, 27 kΩ	C_2, C_3, 100 pF
R_4, 1 MΩ	R_{15}, 4·7 kΩ	C_4, 25 μF
R_5, 33 kΩ	R_{16}, 2·4 MΩ	C_5, C_6, 0·01 μF
R_6, 68 kΩ	R_{17}, 330 Ω	C_7, 0·05 μF
R_7, 22 kΩ	R_{18}, 3·3 kΩ	C_8, 2 μF
R_8, 1 kΩ	R_{19}, 5 kΩ	C_9, C_{10}, 8 μF
R_9, 22 kΩ	R_{20}, 4·7 kΩ	V_1, V_2, CV4003
R_{10}, R_{11}, 100 kΩ	R_{21}, 1 MΩ	V_3, V_4, CV423
R_{12}, 220 kΩ	R_{22}, 47 kΩ	V_5, V_6, CV1832
L, 10 H	T, 350–0–350 V	V_7, CV493

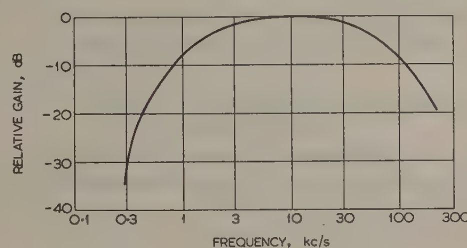


Fig. 2.—Frequency response of counter with 7 m aerial.

circuit in Fig. 1 is shown in Fig. 2. A reasonable cut-off below 1 kc/s has been achieved, but the response extends upwards beyond the nominal upper limit of 20 kc/s.

(5.2) Aerial

The problem of obtaining good response at the lowest frequencies is facilitated if a long aerial is used; however a short aerial is more easily installed, and the 7 m vertical aerial used is a compromise. To avoid possible shielding of the aerial it should not be close to extensive metalwork, such as a steel frame of a building. The short length of feeder, which was not used in the Sullivan and Wells design, was introduced to enable the separation between the aerial and the building housing the counter to be increased.

The aerial coupling capacitor, C_2, should be of good quality, otherwise charge may leak through it and establish a voltage on the grid of V_1 when the aerial is in high static fields. Some observed differences in performance between counters may have been caused by imperfect capacitors.

(5.3) Rectifier, Triggered and Register Circuits

A full-wave rectifier is included to produce a unidirectional voltage for triggering the multivibrator, irrespective of the waveform of the initiating atmospheric.

To minimize the dependence of operation on the duration of the atmospheric, a triggered circuit is included to actuate the register. A hard-valve multivibrator has been used, since thyratrons have been found to be variable in performance.

In the Sullivan and Wells (1957) design a relay was connected in one anode circuit of the multivibrator and this operated the

register. The high current in the register circuit tended to react on the input circuit and led to instability. Although the trouble could be overcome by the addition of decoupling components, use is now made of a register which operates with a small current and can be connected directly in the anode circuit of the multivibrator.

(5.4) Calibration and Automatic Recording

The counter is calibrated by applying a step voltage from a capacitor, the voltage being read on a built-in meter.

A network is included in the cathode circuit of the multivibrator to enable the counts to be marked on a recording milliammeter. The occurrence of a count is indicated by a momentary deflection of the pen by about 1 cm on a 1 mA recorder.

A recorder designed originally for use with a rain gauge has also been utilized. This shows by a pen deflection the total number of counts in predetermined intervals of 1 or 3 min. The pen returns to zero at the end of each interval, while the chart simultaneously moves slightly along the co-ordinate representing time. This recorder is actuated by the closing of a pair of contacts, so a relay must be incorporated in the multivibrator circuit of the counter.

The type of recorder described by Foldès[12] could no doubt also be adapted for use with the recommended design of counter.

(6) SOME MEASUREMENTS OF INSTRUMENTAL PERFORMANCE

The data to be discussed later were obtained during a period when the instruments were still under development and had not been fully standardized. It is therefore necessary to discuss how differences in the design and method of use can affect the results.

(6.1) Aerial Design

The design of the aerial is a critical factor in determining the counter sensitivity. Many comparisons have been made between 7 and 10 m vertical aerials; from an analysis of the circuit the pick-up factor of the longer aerial, including the effects of the feeder, should be about 1·8 times that of the shorter one. The effect of this difference on the relative counts depends on the distribution of storms and on the way in which field strength varies with distance, but the results on various storms conform broadly with the estimated difference in pick-up factor. It should be noted, however, that small differences in aerial length can lead to large differences in counts when storms are distant.

Differences have been noticed between results with aerials of the same length but with the lower ends at different heights above the ground. These were probably caused mainly by the shielding effects of adjacent steel-framed buildings, but it has been considered desirable to standardize the height of the lower end of the aerial at 1·5 m.

Unless otherwise stated, data quoted refer to counters wit 7 m aerials.

(6.2) Comparison between Nominally Identical Counters

Experience in the operation of nominally identical counter side by side, shows that statistically they yield comparabl records over a long period. In the summer of 1957, during a periods with local storms, two such counters registered 942 an 1 068 flashes respectively. Similar counts are usually obtaine on an hourly basis, and even on a 10 min basis, as can be see by the comparisons between two Slough counters shown i Fig. 8.

A flash-by-flash comparison was made for one day with loca storms. Totals on two counters were 294 and 319, but onl 183 were recorded on both counters and the rest did not corre late. The differences were attributed to minor differences in th aerials and to shielding effects. In later stages of the wor improved correlation has been obtained by greater care in aeria standardization, but in comparing data from similar counters i different locations it is unwise to assume that perfect correlatio would have been achieved if they had been on the same site.

(6.3) Comparison between Counters of Different Threshold Settings

A comparison has been made between total counts on instru ments with 3- and 10-volt settings for July and August, 1957 the counts were 1 868 and 823 respectively—a ratio of 2·3 : 1 This small ratio shows that, when integrated statistical result with near storms are considered, the threshold settings are no critical, and suggests also that the range of a 3-volt counter i not greatly in excess of that for 10 volts; in other words, th field strength falls off rather rapidly with increase in distanc A ratio of 2·3 : 1 would nearly correspond to an inverse cub law if the distribution of flashes were uniform.

It should be noted that these observations relate to period with storms at short ranges. Results discussed in Section 8. suggest that, more generally, the field strength varies less rapidl than the inverse cube of the distance.

It will be shown later that differences in threshold setting ca lead to large differences in counts on distant storms.

(7) RELATIONSHIP BETWEEN COUNTER RECORDS AND THUNDER HEARD

In a survey of counter data taken at Slough during 1956 an 1957 it was noticed that local thunder was generally accompanie by an increase in the count rate to at least 30 per hour on 3-volt counter with a 7 m aerial; the correlation is shown i Table 1. The thunder heard is based on Slough observation as far as possible, but since observers were there for only abou 40 hours per week, use has been made also of records fro nearby meteorological stations.

There is a broad similarity between the thunderstorm day

Table 1

Correlation between High Count Rates and Thunder Heard

	Month											
	J	F	M	A	M	J	J	A	S	O	N	D
Days on which count exceeded 30 per hour, 1956 ..								2	2	2	0	0
Thunder heard at or near Slough, 1956								3	3	3	0	0
Days on which count exceeded 30 per hour, 1957 ..	0	0	0	0	1	2	4	4	0	0	2	0
Thunder heard at or near Slough, 1957	1	0	0	0	1	2	2	2	0	0	0	0
Average thunderstorm days from meteorological records	<½	<½	1	1	3	3	3	3	1	1	<½	<½

and the days on which the count exceeded 30 per hour. From May to October in both years thunder was heard on 16 days and the count exceeded 30 per hour on 12 of these. Conversely, the count exceeded 30 per hour on 17 days and thunder was heard on 12 of these.

The two occasions in November, 1957, when the count exceeded 30 per hour were in squally weather. Clouds passing immediately overhead appeared to cause corona discharge from the aerial for a period of a few minutes, but no thunder was heard. Because corona discharge seems to be associated only with thunderclouds passing directly overhead, it is likely to be rather rare at any one station, at least in temperate regions. If continuous recording of count rates is adopted, it can be distinguished by the character of the records.

Mention should be made of a few nights in December, 1957, when there were high counts, although there was no evidence of local storms. The high counts were apparently caused by extensive thunderstorms at ranges of 300–500 km, coupled with good night-time radio propagation. Although the count never exceeded 30 per hour, it exceeded 20 on several occasions. The records were different from those of local storms, showing a slow increase over several hours followed by a decrease over a similar period. Local storms normally cause sudden increases in count rate lasting for only one or two hours.

It would be interesting to know whether the same relationship between count rates and thunderstorm days is valid for other parts of the world, particularly those with high incidence of thunderstorms. Some records have been made in Singapore, Malaya, and Kumasi, Ghana, but unfortunately regular hourly readings could not be taken. However, with some assumptions a useful comparison with the Slough data can be made.

Fig. 3 shows a plot of the peak value of the hourly count on a 3-volt counter at Slough, for each day, against the total count for the day; the plot covers the period July and August, 1957. There is a reasonably close relationship between the two parameters, and a maximum hourly count of 30 corresponds to a daily count of 100. A correlation might therefore be expected between daily counts greater than 100 and thunderstorm days, certainly in England and possibly in other countries.

Fig. 4 compares the number of days in each month with total counts greater than 100 and the thunderstorm days at Singapore and Kumasi. The experimental arrangements

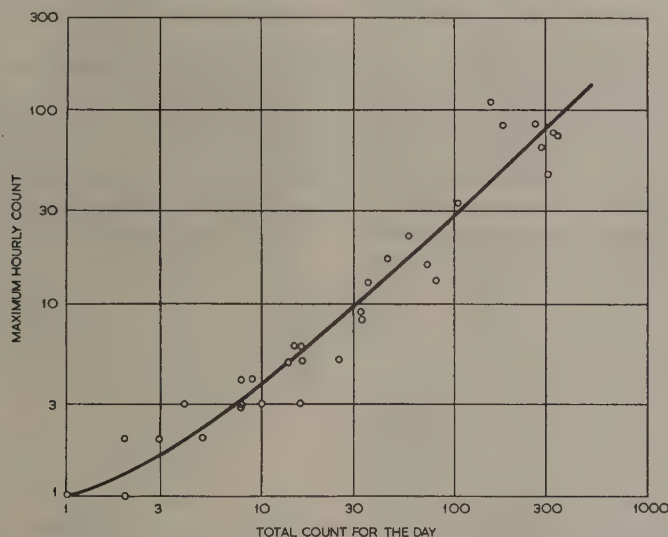


Fig. 3.—Relationship between maximum hourly count and total daily count at Slough.

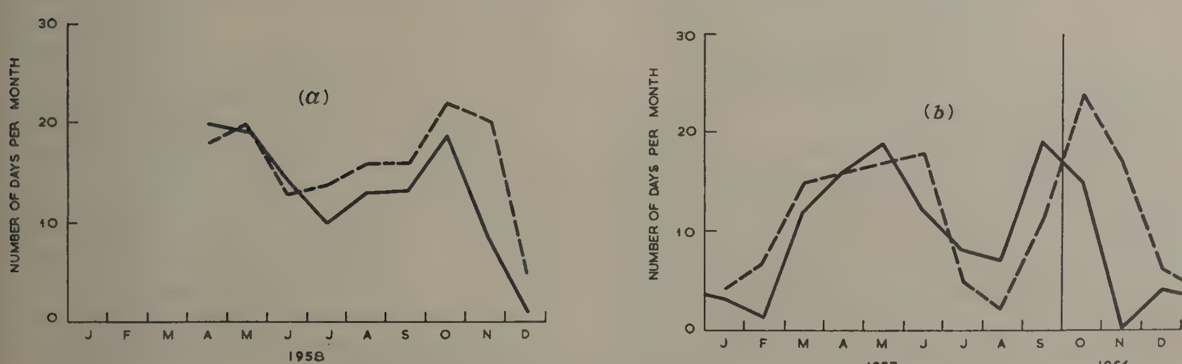


Fig. 4.—Relationship between daily counts and thunderstorm days.

—— Days on which total count exceeded 100.
- - - - Thunderstorm days per month.
(a) Singapore.
(b) Kusami.

differed somewhat from the standard 7 m aerial and the 3-volt sensitivity, but the differences cannot have a marked effect on the plots in Fig. 4, since on most days when the count exceeded 100, it did so by a large factor.

These records confirm that the daily counts can be used as an index of thunderstorm activity roughly corresponding to a thunderstorm day and that the appropriate daily total is about 100.

(8) EFFECTIVE RANGE OF A COUNTER

(8.1) Definition of Effective Range

In view of the amplitude variations of flashes, the range of a counter must be defined somewhat arbitrarily in statistical terms, and comparison between different counters has been hampered by the lack of an agreed method of specifying the range. For the noise application the information required is the average frequency of occurrence of flashes per unit area of the earth's surface. It is suggested that the effective range of a counter be defined as the range within which the actual number of flashes occurring, over a long period, is equal to the number counted. Some of those counted will be outside the effective range, and some within this range will be missed, but the correct average density of the flashes will be deduced.

(8.2) Amplitude Distributions of Atmospherics

For the calculation of effective range from experimental data a knowledge of the amplitude variations from a localized source is required. Several measurements of amplitude distributions from individual storms have been made, sometimes in wide and sometimes in narrow bandwidths. It is found[19, 20] that they can be represented approximately by a log-normal law, i.e. the amplitudes, expressed in logarithmic units, have a normal distribution. The standard deviation can be used as a measure of the variations and some values are given in Table 2.

Table 2

STANDARD DEVIATIONS OF AMPLITUDES OF ATMOSPHERICS FROM LOCALIZED SOURCES

Reference	Type of information	Standard deviation
		dB
Munro *et al.*[19] ..	Narrow band at 100 kc/s; storms at 20–100 miles, in Australia	4
Pierce[5]	Wide band v.l.f.; storms at 100 km, in England	7
Horner[20]	Narrow band at 10 kc/s;	
	(*a*) Storm at 500 km, observed in England	12
	(*b*) Tropical storm at 3 000 km, observed in Australia	8
Storm of 12th August, 1957, discussed in Section 8.5	Narrow band at 6 kc/s; storm at 8 km in England	6

The higher values in Table 2 are for the greater distances, and probably include some variability caused by changing propagation conditions.

The little available information on the variations in intensity from one storm to another, in any one part of the world, suggests that these are normally not large, but more statistical data are required.

(8.3) Techniques for assessing Effective Range

A rough estimate of the effective range may be derived from consideration of the means of calibration. A step-function voltage, normally of 3 volts, is applied, and with a 7 m aerial this is approximately equivalent to a field strength of 3 volts/m. It is known from the work of Appleton and Chapman[21] that a lightning flash produces a field strength which approximates to a step function of 3 volts/m at a distance of 60 km, so the effective range of the counter is likely to be of this order.

The straightforward method of assessing the effective range is to record the positions of all flashes in the area around the counter, by visual observations, direction-finding, etc., and so to determine the range within which the number is equal to the number counted. Although this method is simple in theory, the recording of precise locations of all flashes would be a major undertaking and the observations would need to be continued over a long period to ensure statistical uniformity of the distributions.

Some information can be obtained by studying correlations between the counts on a network of counters, even if the locations of the flashes are not known. For example, if all flashes were of the same intensity, the correlation between individual counts at two stations would decrease almost uniformly from 100% at zero separation to zero at a separation of twice the effective range. The relationship is less simple when the variations in amplitude at the source are considered, and the lack of complete correlation between nominally identical counters at the same place is an added complication. Nevertheless, an approximate indication of the effective range can be obtained by this method, although it requires a rather tedious comparison between records.

It would be advantageous if the effective range could be deduced from observations on a few selected storms at known distances. This can be done if it can be assumed that the observed storms are typical and that over a long enough period the geographical distribution would be uniform. Suppose, for example, that a storm radiates atmospherics with a log-normal amplitude distribution of standard deviation σ decibels, and that the field strength (in volts per metre) is proportional to $(\text{distance})^{-m}$. Let the percentage of the flashes recorded at a distance D from the storm be P. If R is the effective range, Fig. 5 shows how the ratio D/R is related to P for two values of σ (6 and 12 dB) and of m (1 and 3). With the logarithmic and

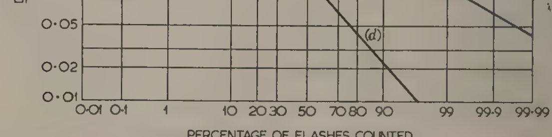


Fig. 5.—Theoretical relationships between percentage count and distance of storm.

(*a*) σ = 6 dB, m = 3.
(*b*) σ = 12 dB, m = 3.
(*c*) σ = 6 dB, m = 1.
(*d*) σ = 12 dB, m = 1.

normal distribution scales used in Fig. 5 the plots are linear, for reasons discussed in Section 12, and the following relationship applies:

$$20 \log_{10} \frac{\text{Distance for } 50\% \text{ count}}{\text{Distance for } 84\% \text{ count}} = \frac{\sigma}{m}$$

If the amplitude variations and propagation law are known, observation of the percentage count from a storm at a known distance gives the effective range. This is derived with least uncertainty if the standard deviation is small, if an inverse-cube propagation law is valid and if the observations are made in conditions where P is about 25%, which is approximately the count at the effective range. It will be shown later (Section 8.6) that observations with counters of different sensitivities at one place can, in theory, be used to derive the values of σ and m.

Reference has been made to the possibility of large differences in counts on distant storms arising from small differences in sensitivity such as might occur with changes in aerial length or threshold setting. With a log-normal amplitude distribution, the counts, when small, are critically dependent on sensitivity. For example, if the standard deviation is 6 dB, a reduction in sensitivity of 3 dB could, with a distant storm, reduce the percentage count from 1 to 0·2—a factor of 5. At a shorter distance a count of 50% would, correspondingly, be reduced only to 31%—a relatively much smaller change. The relationship between sensitivity and counts on individual storms is therefore complex and no simple correction factor can be used to allow for differences between instruments.

(8.4) Probable Locations of Individual Flashes

A further problem which may arise is that of defining the probability that a recorded flash or group of flashes occurred within a given range. This information might be useful to meteorologists in studying the movements of thunderstorms. A theoretical solution to this problem can be found if uniform spatial distribution and amplitude variations are assumed, as before, when a long period is considered. Fig. 6 shows the

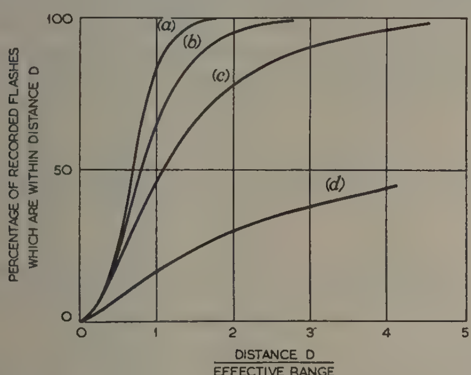


Fig. 6.—Probability of a given recorded flash lying within a given range.

(a) $\sigma = 6$ dB, m = 3. (c) $\sigma = 6$ dB, m = 1.
(b) $\sigma = 12$ dB, m = 3. (d) $\sigma = 12$ dB, m = 1.

probability that a recorded flash occurred within a given distance, related to the effective range, when different amplitude fluctuations and propagation laws are assumed. Except with the lowest curve, the probable distance of the flash is defined with useful accuracy.

It is likely that, by studying the statistics of the incidence of storms and the number of atmospherics they produce, the probable distance of a group of recorded flashes, known to be in the same storm, could be derived with greater accuracy than that of a single flash.

(8.5) Data on the Effective Range of Counters

In the course of investigations into the design and performance of counters, a network has been gradually built up in the London

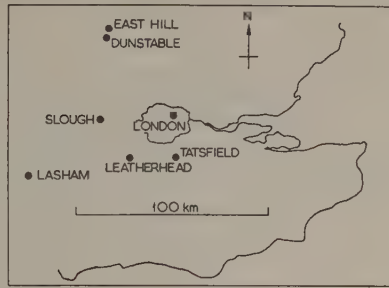


Fig. 7.—Network of counter stations.

area and the present arrangement is shown in Fig. 7. Hourly readings are taken, and in addition, pen records are made at some stations, so enabling correlation between individual flashes to be made.

The techniques for assessing the effective range, as described above, require that there be either a statistically uniform distribution of flashes or a localized storm at a known distance, the atmospherics from which can be positively identified. Unfortunately, storm conditions seldom comply with these requirements. Statistical data are being accumulated but require a long time for their collection and analysis. On the other hand, there have been occasions at Slough when the storm situations have been known and sufficiently well defined for useful information on effective range to be obtained. Some of the results are described below.

The main object of presenting the experimental data is to illustrate techniques of analysis and interpretation. Although the final results are believed to be approximately correct, observations on more storms are necessary to check that they are representative of normal conditions.

Unless there is a statement to the contrary, results relate to a counter with a 7 m aerial. Most data have been obtained with 3-volt counters, but some results with counters of other sensitivities are given.

5th October, 1956.—Experiments during the afternoon of the 5th October, 1956, are of interest as the first attempts at Slough to relate the performance of the counters to known positions of storms. Some records obtained have been described in another publication,[20] in which it was shown that the nearest main storm centre was about 200 km to the south. From direction-finding records it was estimated that less than 11% of the atmospherics radiated from the southerly storm were counted at Slough. This result indicated that the effective range was less than about 150 km.

The counter sensitivity used was 3 volts, but the aerials in use at that time were 9 m long, compared with the 7 m later adopted as standard. It was estimated that the effective range with a 1 m aerial would have been less than 60 km.

5th–6th July, 1957.—One of the few occasions with night-time storms occurred between 2200 h on the 5th July and 0500 h on the 6th July, 1957. The hourly counts on 3-volt instruments at Slough and Tatsfield showed a similar increase to a maximum at about 0200 h on the 6th July. Thunder was heard in London from midnight to 0300 h. The data were examined to determine whether the similarity in the hourly counts was due to the two stations counting the same flashes or experiencing different storms simultaneously. For this purpose the counts in 10 min intervals were plotted, and are shown in Fig. 8. Two counters at Slough show similar plots, while the Tatsfield data are significantly different. No precise values of effective range can be inferred, but it is evident that in this instance the distance of the recorded

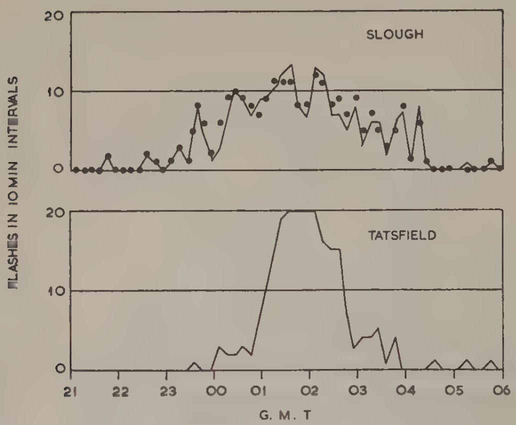


Fig. 8.—Records for 5th–6th July, 1957.

In the Slough data the continuous line is for one counter and the dots are for another.

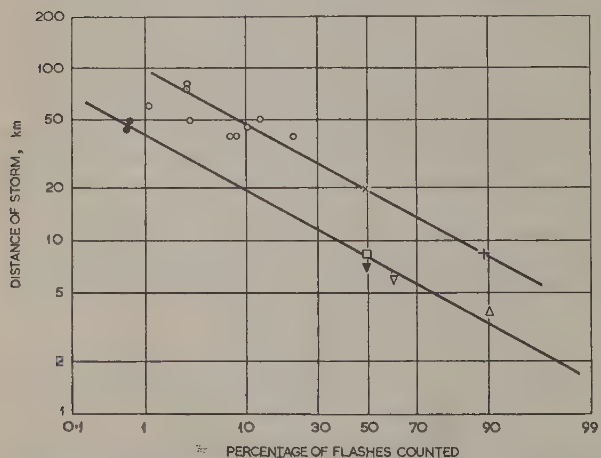


Fig. 9.—Measured percentage counts as a function of distance of storm.

Upper line: 3-volt counter.
Lower line: 10-volt counter.

△ Slough, 10 volts, 19.7.57
+ Slough, 3 volts, 12.8.57
□ Slough, 10 volts, 12.8.57
× Slough, 3 volts, 28.8.58
▼ Slough, 10 volts, 28.8.58
○ Various stations, 3 volts, 5.9.58
● Slough, 10 volts, 5.9.58
▽ Slough, 10 ovlts, 15.9.58

storms was not large compared with the separation of the stations (45 km), or the plots would have been more similar.

19th July, 1957.—During a storm 4 km from Slough, 28 flashes were seen, and of these 21 were seen to strike the ground. Of those seen, 25 triggered a 10-volt counter. In Fig. 9, a 90% count has been plotted at a range of 4 km for comparison with other data, the co-ordinate system being the same as that in Fig. 5.

12th August, 1957.—Of 26 flashes from a storm 8 km from Slough, 23, or 86%, were recorded on a 3-volt counter and 13, or 50%, on a 10-volt counter. These values have been plotted in Fig. 9.

28th August, 1958.—Observations were made on widespread frontal-type storms within a radius of about 50 km from Slough. Although the exact distances of flashes were not known, the data may be used to illustrate a useful relationship between counter records and the amplitudes of atmospherics on a direction-finder tuned to 10 kc/s, with a bandwidth of 300 c/s. Fig. 10 shows the amplitude levels, measured at 10 kc/s, at

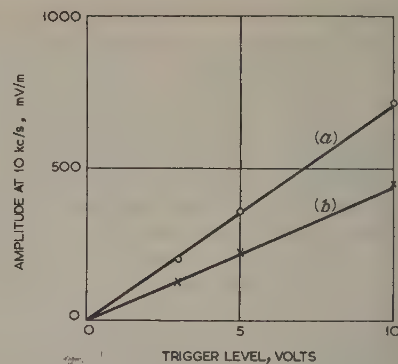


Fig. 10.—Comparison between counter performance and the amplitudes of atmospherics at 10 kc/s.

(*a*) Counter at indicated trigger level counts 50% of atmospherics of stated 10 kc/s amplitude.
(*b*) Corresponding values for 25% count.

which counters of 3-, 5- and 10-volt sensitivities counted 2 and 50% of the flashes. The effective threshold level of 3- an 10-volt counters may therefore be taken as corresponding t 10 kc/s amplitudes of 200 and 700 mV/m respectively, which from past experience[20] are the median amplitudes to be expecte at distances of about 20 km and 7 km. The data therefor suggest that the counter recorded about half the flashes a distances of 20 and 7 km respectively for 3- and 10-volt counters These values have been added to Fig. 9.

5th September, 1958.—The storms of the 5th September, 1958 were of unusual intensity and long duration. They were i existence during the morning near the Isle of Wight and travelle north-eastwards during the day, reaching London shortly afte 1800 h. The data are of particular interest because the number of flashes counted were far greater than any previously recorde by the counters. They are given in detail in Table 3. None o

Table 3

Counter Data for Storms of the 5th September, 1958

G.M.T.	Hourly counts					
	Lasham	Slough	Tatsfield	London	East Hill	Dunstable
10–11	39	0	0	1	0	0
11–12	7	0	0	0	4	0
12–13	31	7	0	4	35	7
13–14	147	6	1	5	18	4
14–15	51	20	3	5	80	28
15–16	(300)	59	7	22	201	35
16–17	(2 000)	(210)	22	61	411	59
17–18	(3 000)	(240)	171	63	449	160
18–19	(2 000)	(340)	2 744	1 690	2 800	936
19–20	(2 000)	260	2 042	773	1 568	262
20–21		45	38	30	783	95
21–22		16	9	36	518	
22–23		8	2	21	339	

Note: The bracketed numbers are approximate estimates deduced from pe records.

the records showed any indication that corona discharge ha occurred.

There is some evidence of differences in sensitivity betwee the counters, these having greatest effect at the low count rates for the reasons given in Section 8.3. The maximum counts ar all of the same order, except those for Dunstable and Slough Dunstable was known to be in operation at a sensitivity o 4 volts instead of 3 volts, and this would reduce the counts t

some extent. At Slough the low counts are believed to be genuine, since only the fringes of the storms were experienced there.

Many of the flashes in these storms were of long duration and some of the near flashes triggered the counter twice or more. The total number of flashes is therefore probably somewhat less than would be deduced from Table 3, but it seems reasonable to assume that the rate of occurrence was of the order of 2000 per hour, at least from 1600 to 2000 h, after which there was a sudden drop in the count rate at Slough, Tatsfield and London.

On several occasions the distances of the storm area from one or more stations were well-defined, and the percentage counts for each known distance, based on an assumed total of 2000 per hour, are plotted in Fig. 9.

15th September, 1958.—A few flashes were observed visually in a storm 6 km from Slough. Of 8 flashes known to be in this storm, all were recorded on a 3-volt counter and 5 on a 10-volt counter. The latter observation has been included in Fig. 9.

(8.6) Discussion of Data on Effective Range

From the storm of the 5th October, 1956, it was deduced that the effective range of a 3-volt counter with a 9 m aerial was probably less than 150 km, and it was inferred that with a 7 m aerial it would have been below 60 km. The storms of the 5th–6th July, 1957, showed that with a 3-volt counter and a 7 m aerial the effective range was not greatly in excess of 45 km and might be less than this.

The subsequent data are plotted in Fig. 9, and on the basis of the reasoning leading to Fig. 5, the best straight lines have been drawn on Fig. 9 through the data for 3- and 10-volt counters. There is some scatter in the data, but this is hardly surprising considering the variable intensities of storms, the possible effects of minor differences in sensitivity and the difficulties of identifying the precise source of every flash. It should be borne in mind also that any change in the propagation index, m, with distance would result in a curved plot, but no account of this can be taken with the present sparse data.

Fig. 5 shows that the effective range is likely to be that corresponding to a 25% count, and on this basis the values derived from Fig. 9 are 30 and 13 km respectively for the 3- and 10-volt counters.

It cannot be expected that accurate quantitative conclusions can be drawn from data on so few storms, but the following arguments illustrate a method for deducing additional information. The ratio between the ranges with 3- and 10-volt counters for any given percentage count is $2 \cdot 3$, which implies that the value of m in the propagation law is $1 \cdot 4$. The ratio between the distances for 50 and 84% counts on either curve is $2 \cdot 0$, and hence, using the equation of Section 8.3, the standard deviation of the amplitudes of the atmospherics is $20m \log_{10} 2 \cdot 0$, or 8 dB. This value is in broad agreement with Table 2. The value of m is low compared with that derived in Section 6.3, but as an average value for a range of distances it is not unreasonable on physical grounds.

In Fig. 6 the derived values of m and σ would give a curve lying between the second and third curves. It therefore seems that a single recorded flash can be assumed to have occurred within the effective range with a probability of about 50% and within twice the effective range with a probability of about 90%.

(9) CONCLUSIONS

(9.1) Operation and Performance

The recommended design of counter, when used with care, provides consistent data.

For instruments to have the same overall sensitivity, particular importance is attached to the lengths of the aerial and feeder, and to the use of an unobstructed site.

Differences in sensitivity are most apparent in the performance on distant storms. Minor differences are not important in observations on storms within the approximate range of 50 km for which the counter was designed.

Corona discharge phenomena may occur during overhead storms but are not common at any one station in England. They may be distinguished by the character of the records when continuous recording is adopted.

(9.2) Relationship between Counts and Thunderstorm Days

With a 3-volt counter and a 7 m aerial at Slough there is an approximate correspondence between thunderstorm days and days on which the hourly count exceeds 30 at some time. A similar correspondence is observed between thunderstorm days and days with a total count exceeding 100.

In Singapore and Ghana the minimum daily count corresponding to a thunderstorm day is also approximately 100.

(9.3) Effective Range

A convenient definition of effective range is the range within which the number of flashes occurring over a long period is equal to the number actually counted.

With certain assumptions regarding the amplitude distribution of atmospherics and the laws governing their propagation, estimates of the effective range may be made from observations on isolated storms. In theory at least, it is possible to deduce the required information on distributions and propagation from the observations themselves, provided that they are made simultaneously with counters of different sensitivities.

Preliminary observations on a few storms in southern England yield values of effective range of 30 and 13 km for 3- and 10-volt counters respectively, with a 7 m aerial. It is also deduced that atmospherics at a given distance from the source have a log-normal amplitude distribution with a standard deviation of about 8 dB, and that the field strength is inversely proportional to the distance raised to a power between $1 \cdot 4$ and 3. More observations are required to evaluate this index more precisely and to determine to what extent it is a function of distance.

Estimates of the effective range may also be made from a study of correlations between individual counts at different stations, but the analysis is rather tedious.

When data are considered statistically on a long-term basis there is a 90% probability, approximately, that a given recorded flash occurred within a distance equal to twice the effective range.

The results on the different storms described are reasonably consistent with each other, and no significant differences have yet been observed between heat-type and frontal-type storms.

(9.4) Further Work

The number of storms observed so far is too small for the data to be regarded with confidence as being typical of conditions in England; still less can they be accepted for world-wide application, and similar observations are required in many more places. The techniques of analysis described appear to provide a satisfactory common basis for comparing results.

To facilitate interpretation and comparison of data it is essential that instruments and quality of sites be standardized so far as possible. This does not necessarily imply that constructional details must be identical, but the overall performance should be standardized in respect of such characteristics as aerial effective height, frequency response, maximum counting rate and calibration technique.

(10) ACKNOWLEDGMENTS

This work was carried out as part of the programme of the Radio Research Board of the Department of Scientific and Industrial Research, and the paper is published by permission of the Director of Radio Research.

Mr. C. Clarke carried out experiments on the optimum design of counter. The assistance provided by Mr. K. C. Whittaker, of Kusami, Ghana, and by staff of the Meteorological Office, British Broadcasting Corporation, Royal Aircraft Establishment and Electrical Research Association, who operated counters, is acknowledged.

(11) REFERENCES

(1) BROOKES, C. E. P.: 'Distribution of Thunderstorms over the Globe'', Meteorological Office Geophysical Memoirs and Professional Notes, No. 24, 1925.
(2) 'World Distribution of Thunderstorm Days, Part I', Report WMO/OMM No. 21, World Meteorological Organization, Geneva, 1953.
(3) 'Revision of Atmospheric Radio Noise Data', C.C.I.R. Report No. 65, International Telecommunications Union, Geneva, 1957.
(4) SCHONLAND, B. F. J.: 'The Lightning Discharge', Volume 22 ('Gas Discharges, II') of 'Encyclopedia of Physics', (Springer Verlag, Berlin, 1956).
(5) PIERCE, E. T.: 'The Development of Lightning Discharges', *Quarterly Journal of the Royal Meteorological Society*, 1955, **81**, p. 229.
(6) TANTRY, B. A. P., SRIVASTAVA, R. S., and KHASTGIR, S. R.: 'Waveform Studies of Electric Field Changes during Cloud-to-Cloud Lightning Discharges', *Proceedings of the National Institute of Science, India*, 1957, **23**, Part A, p. 499.
(7) FORREST, J. S.: 'The Determination of the Location and Frequency of Thunderstorms by a Radio Method', *Quarterly Journal of the Royal Meteorological Society*, 1943, **69**, p. 33.
(8) DAVIS, R.: 'High Voltage Research at the National Physical Laboratory', *Journal I.E.E.*, 1946, **93**, Part I, p. 177.
(9) SCHONLAND, B. F. J., and GANE, P. G.: 'A Lightning Warning Device', *Transactions of the South African Institute of Electrical Engineers*, 1947, **38**, p. 119.
(10) GANE, P. G., and SCHONLAND, B. F. J.: 'The Ceraunometer', *Weather*, 1948, **3**, p. 174.
(11) FOLDÈS, G.: 'Le dénombrement automatique des éclairs proches', Notes Préliminaires du Laboratoire National de Radioélectricité, No. 148, 1951.
(12) FOLDÈS, G.: 'Compteur d'éclairs proches', *op. cit.*, No. 179, 1954.
(13) LUGEON, J., and RIEKER, J.: 'Le compteur d'orages Suisse', *Annalen der Schweizerischen Meteorologischen Zentralanstalt*, 1957, Appendix 9.
(14) SULLIVAN, A. W., WELLS, J. D., and DINGER, H. E.: 'Radio Receiver counts Lightning Strokes', *Electronics*, October 1954, **27**, p. 158.
(15) SULLIVAN, A. W., and WELLS, J. D.: 'A Lightning Stroke Counter', *Bulletin of the American Meteorological Society*, 1957, **38**, p. 291.
(16) ITO, K., KATO, T., and IWAI, A.: 'Local Lightning Flash Counter within 20 km', *Proceedings of the Research Institute of Atmospherics, Nagoya University*, 1955, **3**, p. 69.
(17) PIERCE, E. T.: 'Field Changes and Lightning Flash Counters', *Archiv für Meteorologie, Geophysik und Bioklimatologie*, 1956, **9**, p. 78.
(18) 'An Instrument for Counting Local Lightning Flashes', *Bulletin of the World Meteorological Organization*, 1959, **8**, p. 14.
(19) MUNRO, G. H., WEBSTER, H. C., and HIGGS, A. J.: 'Simultaneous Observations of Atmospherics with Cathode Ray Direction-Finders at Toowoomba and Canberra', Australian Radio Research Board Report No. 8, 1935.
(20) HORNER, F.: 'The Relationship between Atmospheric Radio Noise and Lightning Phenomena', *Journal of Atmospheric and Terrestrial Physics*, 1958, **13**, p. 140.
(21) APPLETON, E. V., and CHAPMAN, F. W.: 'On the Nature of Atmospherics, IV', *Proceedings of the Royal Society*, 1937, **158**, p. 1.

(12) APPENDIX: RELATIONSHIP BETWEEN PERCENTAGE COUNTS AND DISTANCE OF STORM

It is assumed that the amplitudes of the atmospherics at a fixed distance d from the source have a log-normal distribution, i.e. the percentage, P, of flashes having amplitudes greater than E_t decibels is

$$P = \frac{100}{\sigma\sqrt{(2\pi)}}\int_{E_t}^{\infty} \exp\left[-\frac{(E-E_m)^2}{2\sigma^2}\right] dE$$

where E_m is the median amplitude and σ is the standard deviation of E, both in decibels.

If E_t is the threshold sensitivity of the counter the above expression represents the percentage count at distance d from the storm.

It is also assumed that the median field strength decreases according to the power law d^{-m} or, with the field in decibels,

$$E_m = K - 20m \log d$$

where K is a constant.

The percentage count is therefore

$$P = \frac{100}{\sigma\sqrt{(2\pi)}}\int_{E_t}^{\infty} \exp\left[-\frac{(E+20m\log d-K)^2}{2\sigma^2}\right] dE$$

The dependence of P on $20m \log d$ is therefore similar to that on E, and since a plot of P on a normal distribution scale against E on a decibel scale gives a straight line, a linear plot will be obtained by plotting P against $\log d$ in the same co-ordinate system. Also, since E has a standard deviation of σ decibels, $20 \log d$ has an effective standard deviation of σ/m, which leads to the relationship given in Section 8.3.

621.3.011.1

The Institution of Electrical Engineers
Paper No. 3012 M
July 1959

DEVELOPMENT OF THE FORMULAE OF ELECTROMAGNETISM IN THE M.K.S. SYSTEM

By P. VIGOUREUX, D.Sc., Associate Member.

(*The paper was first received* 16*th December*, 1958, *and in revised form* 10*th April*, 1959. *It was published in July*, 1959, *and was read before the* MEASUREMENT AND CONTROL SECTION 2*nd February*, *and the* SOUTHERN CENTRE 24*th February*, 1960.)

SUMMARY

The use of metres, kilogrammes, amperes, etc., instead of centimetres, grammes, deca-amperes (the C.G.S. unit of current), etc., need not change the method of teaching electricity and magnetism, for it does not interfere with the notions of magnetic poles and point charges of electricity. A logical method of developing the theory from conventional experiments on the forces between magnets, between magnets and current-carrying loops, and between electric charges, is to proceed from magnetic poles to magnetic shells and from these to loops of current, and then bring in the definition of the ampere to evaluate the magnetic constant; consideration of the forces between electric charges follows and, together with identification of rate of change of charge and electric current, leads to the law of induction, to Maxwell's equations, and to wave propagation, which, in conjunction with the known velocity of light, gives the value of the electric constant. The theory is then extended to material media.

A treatment suitable for beginners is outlined in Section 9.

(1) INTRODUCTION

In the course of discussions and in papers, many have expressed the view that the teaching of electromagnetism, especially to beginners, had been made more difficult by the introduction of the M.K.S. system.[1–7] The difficulty may have arisen through enthusiasts wishing to take the opportunity of the introduction of the M.K.S. system to change and—as they think—reform the methods of teaching, abandoning the notion of magnetic poles and, at first, even of point charges for those of magnetic and electric force and flux. While there are many ways of developing the theory of electromagnetism, it is by no means true that the method used in the older treatises[8,9] and by some modern authors[10,11] is less satisfactory than it used to be because lengths and masses are expressed in metres and kilogrammes instead of centimetres and grammes, or because the magnetic and electric constants of vacuum are not taken to be unity as formerly.

The method given below, based on three sets of rather conventional experiments and on the internationally accepted definition of the ampere, should prove this point; and although the development is not strictly elementary, since for brevity and—it is hoped—elegance vectors are used throughout, there should be no difficulty in adapting it to suit beginners on the lines indicated in Section 9.

(2) EXPERIMENTAL BASIS

The formulae and equations describing electromagnetic phenomena can be derived from the results of three sets of experiments, namely observation of the forces between magnets, of the similarity of behaviour of coils carrying electric current and of magnets, and of the forces between electric charges.

(3) MAGNETS

Experiments with magnetized needles show that a long thin magnetized body behaves as though it had magnetic poles Φ and $-\Phi$ at the ends if the repulsive force between poles Φ and Φ' a distance r apart is proportional to $\Phi\Phi'/r^2$. The constant of proportionality, which depends on the medium in which the observations are made, cannot be determined from these experiments unless the unit of magnetic pole is known, and since the unit pole is defined in terms of unit electric current, the value of the constant must remain indefinite until unit current, i.e. the ampere, is defined. It is convenient to write the constant as $1/4\pi\mu$, because π or 4π is thereby made to appear in formulae dealing with spherical rather than rectangular distributions. As indicated in Section 8, the value of μ for air differs very little from the value for vacuum and is approximately $1{\cdot}257 \times 10^{-6}$. Returning to the general case, we have

$$F = \frac{\Phi\Phi' \boldsymbol{r}}{4\pi\mu r^3} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

It is convenient to define a vector 'magnetic field strength', $\boldsymbol{H}$, by

$$\boldsymbol{H} = -\nabla\phi \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

where ϕ is a scalar magnetic potential given, in the case of a pole Φ a distance r away, by

$$\phi = \frac{\Phi}{4\pi\mu r} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

The force on a pole Φ' in a field $\boldsymbol{H}$ is

$$\boldsymbol{F} = \Phi'\boldsymbol{H} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

With the definition of $\boldsymbol{H}$ above it can be shown by a purely geometrical proof that

$$\int \mu\boldsymbol{H} \, . \, d\boldsymbol{a} = \Sigma\Phi \quad . \quad . \quad . \quad . \quad . \quad (5)$$

the integral being taken over any surface enclosing all the magnets, and the element of area $d\boldsymbol{a}$ being measured outward. Since observation shows that magnetic poles never occur singly, but in pairs of opposite sign, and that all magnetic material is composed of particles or 'doublets' having equal and opposite poles, in all cases $\Sigma\Phi$ vanishes; moreover, since the left-hand member can be transformed by Green's theorem to $\int \mu\nabla \, . \, \boldsymbol{H}dv$, it is deduced that everywhere

$$\nabla \, . \, \boldsymbol{H} = 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

The potential at $x'y'z'$ (Fig. 1) of a doublet of magnetic

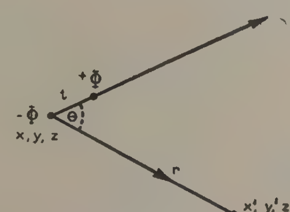


Fig. 1.—Potential of magnetic particle.

Dr. Vigoureux is at the National Physical Laboratory.

moment m or Φl, where l tends to zero while m remains finite, is

$$\phi = \frac{\partial}{\partial l}\left(\frac{\Phi}{4\pi\mu r}\right)l = -\frac{\Phi l}{4\pi\mu r^2}\frac{\partial r}{\partial l} = \frac{m}{4\pi\mu}\frac{\cos\theta}{r^2} \quad . \quad (7)$$

For a uniform shell a of magnetic moment M per unit area or 'strength' M,

$$\begin{aligned} d\phi &= \frac{M\cos\theta da}{4\pi\mu r^2} \\ &= \frac{M}{4\pi\mu}d\Omega \\ &= \frac{M}{4\pi\mu}\frac{\boldsymbol{r}\,.\,\boldsymbol{da}}{r^3} \\ &= \frac{M}{4\pi\mu}\nabla\frac{1}{r}\,.\,\boldsymbol{da} \quad . \quad . \quad . \quad . \quad (8) \end{aligned}$$

where Ω is the solid angle, Fig. 2, subtended at x', y', z' by the positive face of the shell.

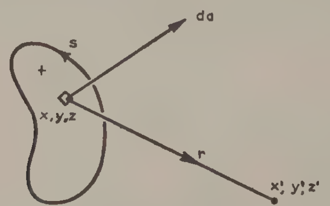


Fig. 2.—Potential of magnetic shell.

Then

$$\phi = \frac{M}{4\pi\mu}\Omega = \frac{M}{4\pi\mu}\int\nabla\frac{1}{r}\,.\,\boldsymbol{da} \quad . \quad . \quad . \quad . \quad (9)$$

The magnetic field strength at x', y', z' is

$$\boldsymbol{H} = -\nabla\phi$$

where the gradient is with respect to x', y', z'. But since the gradient of r with respect to x', y', z' is equal and opposite to that of r with respect to x, y, z,

$$\begin{aligned} \boldsymbol{H} &= -\nabla\phi \text{ (gradients w.r.t. } x', y', z'\text{)} \\ &= \nabla\phi \text{ (gradients w.r.t. } x, y, z\text{)} \\ &= \frac{M}{4\pi\mu}\int\nabla\nabla\frac{1}{r}\,.\,\boldsymbol{da} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (10) \end{aligned}$$

The field strength can be referred to the contour only of the surface a, for by a theorem which can be proved in the same way as Stokes's theorem (see Appendix), if $\nabla\,.\,\boldsymbol{U}$ vanishes,

$$\int_a \nabla\boldsymbol{U}\,.\,\boldsymbol{da} = -\int_s \boldsymbol{U}\times\boldsymbol{ds} \quad . \quad . \quad . \quad . \quad (11)$$

Here $\nabla\,.\,\nabla\frac{1}{r}$ vanishes, so the condition is satisfied, and

$$\boldsymbol{H} = -\frac{M}{4\pi\mu}\int\nabla\frac{1}{r}\times\boldsymbol{ds} \quad . \quad . \quad . \quad . \quad (12)$$

which by expansion into the three components can be shown to be equivalent to

$$\boldsymbol{H} = -\frac{M}{4\pi\mu}\int\nabla\times\frac{\boldsymbol{ds}}{r} \quad . \quad . \quad . \quad . \quad (13)$$

The form of this expression verifies that $\nabla\,.\,\boldsymbol{H}$ always vanishes, and it suggests that $\boldsymbol{H}$ can be expressed in terms of a 'vector magnetic potential' $\boldsymbol{A}$ defined by

$$\nabla\times\boldsymbol{A} = \mu\boldsymbol{H} \quad . \quad . \quad . \quad . \quad . \quad . \quad (14)$$

Remembering that in this formula giving $\boldsymbol{H}$ at the point x', y', z' the differentiations are effected with respect to x', y', z', whereas in the integral expression for $\boldsymbol{H}$ above they are with respect to x, y, z and therefore of opposite sign, we find

$$\boldsymbol{A} = \frac{M}{4\pi}\int\frac{\boldsymbol{ds}}{r} = -\frac{M}{4\pi}\int\nabla\frac{1}{r}\times\boldsymbol{da} \quad . \quad . \quad . \quad (15)$$

For completeness a last expression for $\boldsymbol{H}$ can be given in terms of the surface of the shell, for by another theorem analogous to Stokes's (see Appendix), if $\nabla\times\boldsymbol{U}$ vanishes,

$$\int\boldsymbol{U}\times\boldsymbol{ds} = -\int\nabla\times\boldsymbol{U}\times\boldsymbol{da} \quad . \quad . \quad . \quad (16)$$

Here $\boldsymbol{U}$ is $\nabla\frac{1}{r}$ so its curl vanishes, the condition is satisfied, and eqn. (12) becomes

$$\boldsymbol{H} = \frac{M}{4\pi\mu}\int\nabla\times\nabla\frac{1}{r}\times\boldsymbol{da} \quad . \quad . \quad . \quad (17)$$

We now suppose that x', y', z' is a point on a shell a' (Fig. 3), and calculate the mutual potential energy, W, of the shells.

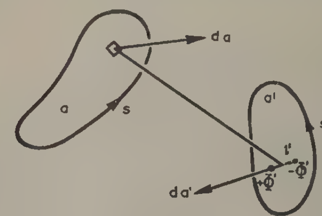


Fig. 3.—Mutual energy of magnetic shells.

The work done to bring a particle $\boldsymbol{m}'$ or $\Phi'\boldsymbol{l}'$ from infinity to x', y', z' is

$$dW = \frac{\partial}{\partial l'}(\Phi'\phi)l' = \boldsymbol{m}'\,.\,\nabla\phi = -\boldsymbol{m}'\,.\,\boldsymbol{H} \quad . \quad . \quad (18)$$

where ϕ and $\boldsymbol{H}$ are the potential and field of a; so, for the whole shell a',

$$\begin{aligned} W &= -M'\int\boldsymbol{H}\,.\,\boldsymbol{da}' \\ &= -\frac{M'}{\mu}\int\nabla\times\boldsymbol{A}\,.\,\boldsymbol{da}' \\ &= -\frac{M'}{\mu}\int\boldsymbol{A}\,.\,\boldsymbol{ds}' \\ &= -\frac{MM'}{4\pi\mu}\int_{s'}\int_s\frac{\boldsymbol{ds}\,.\,\boldsymbol{ds}'}{r} \quad . \quad . \quad . \quad (19) \end{aligned}$$

(4) CURRENTS

In the second set of experiments the behaviour of virtually closed loops of current is compared with that of magnets, and it is found that a current I in a small loop of area da behaves like a magnetic particle and that the equivalent moment of the loop is proportional to $I\boldsymbol{da}$. By extension, a loop s carrying a current I has a strength proportional to I and therefore a

potential proportional to $I\Omega$. Since the unit magnetic pole was left undefined pending the definition of unit current, the constant of proportionality is in this case arbitrary and the simplest choice is to write

$$\phi = \frac{I\Omega}{4\pi} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (20)$$

the 4π being retained for the same reason as before. Thus I takes the place of M/μ in the preceding formulae, and the mutual potential energy of loops s, s' carrying currents I, I' is

$$W = -\frac{\mu I I'}{4\pi}\int\int\frac{ds\,.\,ds'}{r} \quad . \quad . \quad . \quad . \quad (21)$$

which is Neumann's formula, in which μ still remains to be determined from the definition of the ampere.

The mechanical force exerted by one loop on the other is equal to the negative gradient of W, but it is often convenient to calculate the force in two steps by finding first the force due to the loop s on an element ds' of s'. This calculation is most readily effected by taking the negative gradient of eqn. (19) after substitution of $\mu I'$ for M'; thus

$$F = \mu I'\int \nabla H\,.\,da' \text{ (differentiations w.r.t. } x', y', z') \quad (22)$$

Here, as seen before, $\nabla\,.\,H$ vanishes everywhere. But when this condition holds the above expression is, by eqn. (11),

$$F = -\mu I'\int H \times ds' \quad . \quad . \quad . \quad . \quad (23)$$

Thus, provided that integration is eventually carried out all round s', it is permissible to write for the force exerted by the loop s on an element ds' of s'

$$dF = -\mu I' H \times ds' \quad . \quad . \quad . \quad . \quad . \quad . \quad (24)$$

which expresses what is often called Ampère's law.

Substitution of I for M/μ in eqns. (17), (12), (15) and (19) leads to the following expressions for the field H of a current I in a loop s:

$$H = \frac{I}{4\pi}\int \nabla \times \nabla\frac{1}{r} \times da$$

$$= -\frac{I}{4\pi}\int \nabla\frac{1}{r} \times ds$$

$$= -\frac{I}{4\pi}\int\frac{r \times ds}{r^3} \quad . \quad . \quad . \quad . \quad . \quad . \quad (25)$$

$$A = \frac{\mu I}{4\pi}\int\frac{ds}{r} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (26)$$

$$W = -\mu I'\int H\,.\,da' \quad . \quad . \quad . \quad . \quad . \quad . \quad (27)$$

(5) THE AMPERE

The ampere is defined as that current which, maintained in two straight parallel filaments of infinite length one metre apart, produces between them in vacuum a force of 2×10^{-7} newton per metre length. The formulae found above for H and dF can now be applied to the calculation of the force in this case for comparison with the definition.

In Fig. 4, $-r \times ds$ is in the negative direction of x and equal to $rds\cos\theta$, and integration gives

$$H_x = -\frac{2I}{4\pi b} \quad . \quad . \quad . \quad . \quad . \quad . \quad (28)$$

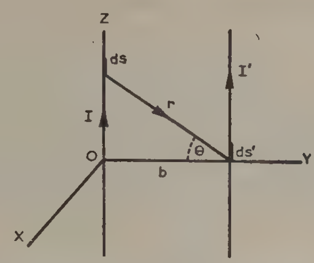


Fig. 4.—Force between straight parallel wires.

Then if $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$ represent unit vectors along X, Y, Z,

$$dF = \frac{2\mu I I'}{4\pi b}(\boldsymbol{i} \times \boldsymbol{k})ds' = -\boldsymbol{j}\frac{2\mu I I'}{4\pi b}ds' \quad . \quad . \quad (29)$$

The force is thus attractive and for a metre length of one of the wires and a metre separation it is $2\mu/4\pi$ which when equated to the force in the definition gives

$$\mu = 4\pi \times 10^{-7} \quad . \quad . \quad . \quad . \quad . \quad . \quad (30)$$

The unit magnetic pole and associated quantities are thereby defined.

(6) ELECTRIC CHARGES

Experiments with electric charges show that the repulsive force between charges Q and Q' a distance r apart is proportional to QQ'/r^2; as for magnets, the constant of proportionality cannot be determined from these experiments alone, and it will be settled when the consequences of the results of the three sets of experiments have been worked out. But, as before and for the same reason, it is written $1/4\pi\epsilon$. Then

$$F = \frac{QQ'r}{4\pi\epsilon r^3} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (31)$$

A vector electric field strength E is defined by

$$E = -\nabla V \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (32)$$

where V is a scalar electric potential given, in the case of a charge Q a distance r away, by

$$V = \frac{Q}{4\pi\epsilon r} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (33)$$

The force on a charge Q' in a field E is

$$F = Q'E \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (34)$$

With the definition of E above, a purely geometrical proof leads to Gauss's theorem:

$$\int \epsilon E\,.\,da = \Sigma Q \quad . \quad . \quad . \quad . \quad . \quad . \quad (35)$$

the integral being taken over any surface enclosing all the charges, and the elements of area da being measured outwards. By extension to charges distributed with volume density ρ and application of Green's theorem in the form

$$\int E\,.\,da = \int \nabla\,.\,E dv \quad . \quad . \quad . \quad . \quad (36)$$

it is found that at any point where the volume density of charge is ρ,

$$\epsilon\nabla\,.\,E = \rho \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (37)$$

Thus, wherever there is no charge,

$$\nabla . E = 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad (38)$$

An extension of the experiments with electric charges and currents, e.g. discharge of condensers through galvanometers, indicates proportionality between Q and I, and it is permissible to choose the unit of charge, or coulomb, such that $Q = I$, and determine the resulting value of ϵ. In vacuum there are no steady currents, but Gauss's theorem leads to the notion, across any element da of a surface near a varying charge, of displacement current $\epsilon \dot{E} . da$.

(7) MAXWELL'S EQUATIONS

The experiments with electric currents will have shown that, in order to maintain a current I in a conducting loop s, some agent, e.g. battery or dynamo, is needed; hence the notion of electromotive force, e, of electrical resistance, R, and of conductivity, σ. But if in order to take a charge Q round the loop at a rate such that $\dot{Q} = I$ an electric field strength E is required, the work which must be done against resistance is $Q \int E . ds$, and the power required is $\dot{Q} \int E . ds$ or $I \int E . ds$. The e.m.f. can be identified with $\int E . ds$ and the power written as

$$P = I \int E . ds = Ie \quad . \quad . \quad . \quad . \quad . \quad (39)$$

The mutual energy of a loop s carrying a current I in a field H is, by eqn. (27),

$$W = - \mu I \int H . da = - IN \quad . \quad . \quad . \quad (40)$$

for brevity.

This formula has been used to deduce, by differentiation and comparison with eqn. (39), the law of induction of e.m.f. in a closed path in which the magnetic flux varies, i.e.

$$e = - \frac{\partial N}{\partial t} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (41)$$

Maxwell[8] reproduces Helmholtz's proof, which is repeated in a slightly modified form by Jeans.[9] Maxwell states, however, that the argument is not satisfactory, because variations which may occur in the currents are neglected. If that is so, it would appear that Faraday's experiments on electromagnetic induction should be added to the three sets of experiments mentioned in Section 2.

Whether eqn. (41) is established by argument or experiment, it can be put into the form

$$\int E . ds = - \mu \int \dot{H} . da \quad . \quad . \quad . \quad . \quad (42)$$

and transforming the left-hand side by Stokes's theorem

$$\int \nabla \times E . da = - \mu \int \dot{H} . da \quad . \quad . \quad . \quad (43)$$

or

$$- \nabla \times E = \mu \dot{H} \quad . \quad . \quad . \quad . \quad . \quad . \quad (44)$$

Again, if the field near a current I is H, work $- \int H . ds$ is done against the field when a unit pole is moved round a closed path s surrounding the current. This work is equal to the increase in potential energy, i.e. to $I/4\pi$ times the increase in the solid angle subtended by the positive face of the current loop. This increase can be seen to be -4π, thus

$$-I = - \int H . ds = - \int \nabla \times H . da \quad . \quad . \quad (45)$$

by Stokes's theorem, where a is any surface bounded by the closed path s and therefore crossed by the current. By adding to I any displacement current there may be and by extending the formula to the case of current distributed over a with density J or σE,

$$\int \nabla \times H . da = \int (J + \epsilon \dot{E}) . da \quad . \quad . \quad . \quad (46)$$

or

$$\nabla \times H = \epsilon \dot{E} + J \quad . \quad . \quad . \quad . \quad . \quad . \quad (47)$$

For the sole purpose of determining ϵ from Maxwell's equations (44) and (47), it is permissible to consider a region of space free of electric charges and of conductors; the equations then reduce to

$$\nabla \times H = \epsilon \dot{E} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (48)$$

$$-\nabla \times E = \mu \dot{H} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (49)$$

which combine to give

$$\nabla \times \nabla \times E + \epsilon \mu \ddot{E} = 0 \quad . \quad . \quad . \quad . \quad (50)$$

or

$$\nabla \nabla . E - \nabla . \nabla E + \epsilon \mu \ddot{E} = 0 \quad . \quad . \quad . \quad (51)$$

and as in a region where there are no space charges the first term vanishes,

$$\nabla . \nabla E = \epsilon \mu \ddot{E} \quad . \quad . \quad . \quad . \quad . \quad . \quad (52)$$

which points to wave propagation with velocity $1/\sqrt{(\epsilon\mu)}$ determined experimentally to be equal to $2 \cdot 997925 \times 10^8$m/s, from which, with the value of μ above, ϵ is $8 \cdot 854173 \times 10^{-12}$.

It is usual to write for convenience

$$\epsilon E = D \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (53)$$

$$\mu H = B \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (54)$$

and to call D and B densities of electric and magnetic flux. Maxwell's equations then become

$$\nabla \times H = \dot{D} + J \quad . \quad . \quad . \quad . \quad . \quad . \quad (55)$$

$$-\nabla \times E = \dot{B} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (56)$$

With the definition of the vector potential A given before, and the present notation,

$$B = \nabla \times A \quad . \quad . \quad . \quad . \quad . \quad . \quad (57)$$

and if the formula (26) for A be extended to currents distributed with density J,

$$A = \frac{\mu}{4\pi} \int \frac{J dv}{r} \quad . \quad . \quad . \quad . \quad . \quad . \quad (58)$$

taken over the volume containing the currents, differentiation to find B being with respect to the point at which A is evaluated.

A corresponding extension of eqn. (33) gives for the scalar electric potential

$$V = \frac{1}{4\pi\epsilon} \int \frac{\rho dv}{r} \quad . \quad . \quad . \quad . \quad . \quad . \quad (59)$$

where ρ is density of electric charge.

If V is the potential due to charges only, then in the absence of currents and of magnets,

$$E = - \nabla V \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (60)$$

But since, in general,

$$-\nabla \times E = \dot{B} = \nabla \times \dot{A} \quad . \quad . \quad . \quad . \quad (61)$$

a solution is

$$E = - \dot{A} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (62)$$

and the general solution

$$E = - \dot{A} - \nabla V \quad . \quad . \quad . \quad . \quad . \quad . \quad (63)$$

(8) MATERIAL MEDIA

The results and formulae hitherto given have been obtained on the assumption that the experiments from which they were deduced had been made in vacuum. In practice, except for the determination of the velocity of propagation of electromagnetic waves, they will have been made in air, but the assumption is justified because selected experiments show that the difference between air and vacuum for this purpose is so small as to be almost negligible. There are substances, however, for which the electric and magnetic effects, although similar in nature to those observed in vacuum or in air, differ considerably in magnitude. For example, if the inside of a closely wound toroidal coil of wire carrying current is filled with iron, the magnetic flux density there, as determined for instance from the electromotive force induced in an auxiliary winding on reversal of the current, is many times the value for air; the ratio of the flux densities with and without iron is called the relative permeability, or simply permeability if there is no risk of ambiguity, of the iron. Similarly, if the space between the plates of a condenser is filled with some insulator (gaseous, liquid or solid), the same voltage applied to the plates is found to give rise to a larger charge; the ratio of the charges with and without the insulator is called the relative permittivity or simply the permittivity of the insulator.

It can be deduced from observations of this sort that the formulae given in the preceding Sections apply also when space is filled by some 'medium' other than vacuum, provided that the magnetic and electric constants of vacuum, μ and ϵ, are multiplied by the permeability and permittivity respectively. It is normal to use μ and ϵ for those products and to call the values for vacuum μ_0 and ϵ_0.

Although the formulae can be deduced from observations like those described above, their direct verification is not easy. Measurement of the force between magnets or between currents when the medium is a solid is complicated; the relative permeability of gases is too close to unity for comparison of forces in a gas and in vacuum to be significant, and the writer is not aware that even with magnetic liquids, e.g. solutions of certain salts of iron, there is direct experimental proof that the forces between magnets are inversely proportional to the permeability, whereas the forces between currents are directly proportional to it. Verification of the effect of permittivity on force between electric charges is perhaps easier, although still troublesome. It may be, however, that there has been little effort at direct verification of the laws of force, because indirect verification depending on the application of other formulae derived from those laws has been ample for electricity as well as for magnetism.

(9) ELEMENTARY TREATMENT

Since some writers[3, 6] have expressed the view that the M.K.S. system, although possibly convenient for advanced courses, is more difficult than the C.G.S. systems for beginners, an outline of what seems a reasonable procedure may not be out of place here. The beginner is supposed to be familiar with trigonometry, but not with the calculus.

Experiments with magnets lead as before to an expression of the form $k\Phi\Phi'/r^2$, where k is a constant. Here and now we explain that, for reasons of convenience which will become clear as we proceed, we write k in the form $1/4\pi\mu$; and further that, with the units in which magnetic quantities are normally expressed, the value of μ is $1{\cdot}257 \times 10^{-6}$.

The notions of magnetic potential, magnetic field strength, magnetic flux and flux density follow in the usual elementary way. Indeed, with a unit pole giving rise to unit flux a simple notion of Gauss's theorem does not seem excluded, even for beginners, nor do those of magnetic doublet, magnetic moment and magnetic shell. The potential due to a doublet, that due to an element of shell, and finally expression of the potential of a shell in terms of its strength and of the solid angle subtended, can all be derived without the use of the calculus.

The second series of experiments demonstrating the similarity of behaviour of coils carrying electric currents and of magnets must be extended to illustrate in special cases the action of a magnetic field on a current and the magnetic force produced by a current. It does not seem possible to explain to beginners not conversant with the calculus how to deduce the law

$$H = \frac{I \sin \theta ds}{4\pi r^2} \quad . \quad . \quad . \quad . \quad . \quad (64)$$

from the formula for the potential of a magnetic shell. The best that can be done is to show that, when this law is applied to simple cases, it agrees with results obtained in a more elementary way, e.g. for the field strength at the centre or on the axis of a circle. The beginner can also be shown how to calculate the field strength inside an infinite solenoid or near an infinitely long straight wire from the expression for the potential of a shell, but he would not be able to apply formula (64) to these cases.

The corresponding expression

$$F = \mu IH \sin \theta ds \quad . \quad . \quad . \quad . \quad . \quad (65)$$

can be obtained from the first by simple processes, and so afterwards can the force per unit length between two parallel wires. The definition of the ampere is now introduced and provides a check of the value stated before for μ.

The procedure for electric charges is similar to that for magnetic poles. Here again it is necessary to introduce a constant $1/4\pi\epsilon$ and to say that with the unit of charge in common use ϵ is $8{\cdot}85 \times 10^{-12}$. The notions of electric field strength and electric potential follow as before, and lead to experiments with condensers, discharge through galvanometers and identification of rate of change of charge with electric current. Finally electromagnetic induction must be demonstrated by experiments with magnets and electric circuits.

With the limitations put on the permissible level of mathematics it does not seem possible to go much further. Maxwell's equations and wave propagation cannot be treated without the use of the calculus, and in consequence the beginner cannot be shown how to check the value of ϵ. But this limitation is precisely the same as that with which he is confronted in the C.G.S. systems, where he has to accept that the ratio of the unit of electric charge in the electromagnetic system to the unit of electric charge in the electrostatic system is equal to the velocity of light. Is it any worse to ask him to accept that that velocity is equal to $1/\sqrt{(\epsilon\mu)}$?

(10) CONCLUSION

The use of metres, kilogrammes, amperes, etc., instead of centimetres, grammes, deca-amperes (the C.G.S. unit of current), etc., need not change the method of teaching electricity and magnetism. The old method is still adequate; in fact, it becomes simpler than it used to be if the units of all quantities encountered are based on the internationally agreed definition of the ampere. It is not suggested that it is the only method or the only good method of approaching the subject, but that it is a good one.

(11) REFERENCES

(1) Carr, L. H. A., *et al.*: 'Symposium of Papers on the M.K.S. System of Units', *Proceedings I.E.E.*, 1950, **97**, Part I, p. 235.
(2) Gibbs, P. C., *et al.*: 'M.K.S. Units for School Physics?', *Journal I.E.E.*, 1956, **2**, p. 152.

(3) LIPSON, H.: 'The M.K.S. System or the Teaching of Elementary Electricity and Magnetism', *Bulletin of the Institute of Physics*, 1957, **8**, p. 13.

(4) WESTON, J. D.: 'Teaching Electromagnetic Theory', *ibid.*, p. 317.

(5) HERCUS, E. O.: 'The M.K.S.A. System and Ampère's Electrodynamics', *ibid.*, p. 388.

(6) MACFADYEN, K. A.: 'M.K.S. Units in Physics Teaching', *ibid.*, 1958, **9**, p. 34.

(7) SILLITTO, R. M.: 'M.K.S. Units in Physics Teaching', *ibid.*, p. 127.

(8) MAXWELL, J. C.: 'A Treatise on Electricity and Magnetism' (Clarendon Press, 1873).

(9) JEANS, J. H.: 'The Mathematical Theory of Electricity and Magnetism' (Cambridge University Press, 1927).

(10) ASTBURY, N. F.: 'Introduction to Electrical Applied Physics' (Chapman and Hall, 1956).

(11) BLEANEY, B. I., and BLEANEY, B.: 'Electricity and Magnetism' (Clarendon Press, 1957).

(12) ASTBURY, N. F.: 'M.K.S. Units in Physics Teaching', *Bulletin of the Institute of Physics*, 1958, **9**, p. 128.

(13) HAMMOND, P.: 'A Short Modern Review of Fundamental Electromagnetic Theory', *Proceedings I.E.E.*, Paper No. 1595, December, 1953 (**101**, Part I, p. 147).

(14) REKVELD, J.: 'New Approach to the Teaching of Electrostatics in General Physics Courses', *American Journal of Physics*, 1958, **26**, p. 139.

(12) APPENDIX: FORMULAE CONNECTING LINE AND SURFACE INTEGRALS

Two theorems have been used in Section 3 for transforming surface into line integrals. In view if their simplicity and usefulness it would be surprising if they had not been enunciated and applied before. However, since no references are known to the author, the proof is given here for the sake of completeness.

Theorem.—The integral $\int \boldsymbol{U} \times \boldsymbol{ds}$ taken round the boundary s of a surface a is equal to $-\int \nabla \boldsymbol{U} . \boldsymbol{da}$ if $\nabla . \boldsymbol{U}$ vanishes, and it is equal to $-\int \nabla \times \boldsymbol{U} \times \boldsymbol{da}$ if $\nabla \times \boldsymbol{U}$ vanishes.

Divide the surface into triangles so small that their sides may be considered straight. It is sufficient to prove the theorem for one such triangle, because the sum of the surface integrals over all the triangles is equal to the integral over the surface a, and the sum of the line integrals is equal to the line integral round the boundary, since every side not part of the boundary is traced twice in opposite senses.

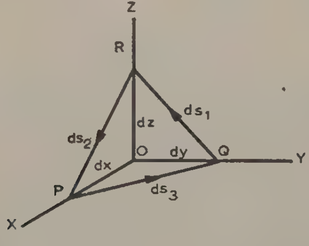


Fig. 5.—Triangular element of surface.

Let PQR (Fig. 5) be one such triangle, and denote by suffixes x, y, z the projections of line elements on the X, Y, Z axes or of surface elements on the X, Y, Z planes. Then

$$\tfrac{1}{2}dydz = da_x, \text{ etc.}$$

Denote by a suffix 0 the value of a quantity at the origin and by the suffix m the mean value over a side of the triangle.

Then: on ds_1, $(U_z)_m = (U_z)_0 + \frac{1}{2}\frac{\partial U_z}{\partial y}dy + \frac{1}{2}\frac{\partial U_z}{\partial z}dz$

on ds_2, $(U_z)_m = (U_z)_0 + \frac{1}{2}\frac{\partial U_z}{\partial z}dz + \frac{1}{2}\frac{\partial U_z}{\partial x}dx$

on ds_3, $(U_z)_m = (U_z)_0 + \frac{1}{2}\frac{\partial U_z}{\partial x}dx + \frac{1}{2}\frac{\partial U_z}{\partial y}dy$

Also $(ds_1)_y = -dy$, $(ds_2)_y = 0$, $(ds_3)_y = dy$.

Thus the sum of the components of $U_z ds_y$ over the triangle i

$$\Sigma U_z ds_y = -(U_z)_0 dy - \frac{1}{2}\frac{\partial U_z}{\partial y}dydy - \frac{1}{2}\frac{\partial U_z}{\partial z}dzdy$$

$$+ (U_z)_0 dy + \frac{1}{2}\frac{\partial U_z}{\partial x}dxdy + \frac{1}{2}\frac{\partial U_z}{\partial y}dydy$$

$$= -\left(\frac{\partial U_z}{\partial z}da_x - \frac{\partial U_z}{\partial x}da_z\right) \quad . \quad . \quad . \quad . \quad . \quad (66$$

Similarly, $\Sigma U_y ds_z = \frac{\partial U_y}{\partial y}da_x - \frac{\partial U_y}{\partial x}da_y \quad . \quad . \quad . \quad (67$

Subtracting eqn. (66) from eqn. (67) gives

$$\Sigma(U_y ds_z - U_z ds_y) = -\frac{\partial U_y}{\partial x}da_y - \frac{\partial U_z}{\partial x}da_z + \frac{\partial U_y}{\partial y}da_x + \frac{\partial U_z}{\partial z}da \quad . \quad . \quad . \quad . \quad (68$$

The right-hand term of this expression can be written

$$-\frac{\partial}{\partial x}(U_x da_x + U_y da_y + U_z da_z) + \left(\frac{\partial U_x}{\partial x} + \frac{\partial U_y}{\partial y} + \frac{\partial U_z}{\partial z}\right)da_x$$

If $\nabla . \boldsymbol{U}$ is zero, only the first term is left; equating it to th left-hand side of eqn. (68) and integrating both sides over a the triangles yields

$$\int (U_y ds_z - U_z ds_y) = -\int \frac{\partial}{\partial x}(U_x da_x + U_y da_y + U_z da_z)$$

which are the x-components of the vector formula

$$\int \boldsymbol{U} \times \boldsymbol{ds} = -\int \nabla \boldsymbol{U} . \boldsymbol{da} \quad . \quad . \quad . \quad . \quad (6$$

But the right-hand term of eqn. (68) can also be put int the form

$$-\frac{\partial}{\partial y}(U_x da_y - U_y da_x) + \frac{\partial}{\partial z}(U_z da_x - U_x da_z)$$

$$+\left(\frac{\partial U_x}{\partial y} - \frac{\partial U_y}{\partial x}\right)da_y - \left(\frac{\partial U_z}{\partial x} - \frac{\partial U_x}{\partial z}\right)da_z \quad . \quad . \quad (7$$

If $\nabla \times \boldsymbol{U}$ is zero, the last two terms vanish, and proceedin as before we find

$$\int (U_y ds_z - U_z ds_y) = -\int\left[\frac{\partial}{\partial y}(U_x da_y - U_y da_x) - \frac{\partial}{\partial z}(U_z da_x - U_x da_z)\right.$$

which are the x-components of the vector formula

$$\int \boldsymbol{U} \times \boldsymbol{ds} = -\int \nabla \times \boldsymbol{U} \times \boldsymbol{da} \quad . \quad . \quad . \quad (7$$

DISCUSSION BEFORE THE MEASUREMENT AND CONTROL SECTION, 2ND FEBRUARY, AND THE SOUTHERN CENTRE AT PORTSMOUTH, 24TH FEBRUARY, 1960

Mr. P. Hammond: Many of the early protagonists of the M.K.S. system confused the question of units with that of physical concept. The author makes a clear distinction between the two matters and shows that the sequence of instruction can be quite independent of the system of units. I am particularly glad to see that he retains the concept of magnetic pole strength, which is essential to a proper understanding of the behaviour of iron in a magnetic field. In teaching I prefer to start with electric charges before dealing with magnetic poles, but otherwise I use a method similar to that of the author.

I have, however, strong misgivings about the Section dealing with material media. So far as I am aware it is not true that the inverse square law of electricity or magnetism depends on the medium in which it acts: like the inverse square law of gravity, it is unaffected by the medium. Additional forces come into play if the medium is polarizable, but it is misleading to the student to lump all these forces together and to say that the forces are reduced in a constant ratio. I should like the author to attempt the discussion of permanent magnets in terms of relative permeability: he will find it very difficult.

The second point which worries me is the author's treatment of Maxwell's equations and particularly the expressions for the potentials [eqns. (58) and (59)]. Eqn. (58) for the vector potential has been derived by making use of a scalar magnetic potential. This implies the absence of displacement current. The entire mechanism of electromagnetic radiation depends on the use of the Lorenz retarded potentials.

My last point concerns eqn. (64), which to the author seems to need the calculus for its derivation. I derive this useful formula without the calculus by considering the work done in crossing a magnetic shell. I then apply this 'work law' to the Heaviside rational current-element. This procedure avoids the use of the magnetic potential and the difficult geometrical concept of solid angle.

Mr. C. J. Carpenter: I agree entirely that the method of presentation need not—and should not—be affected by the units used, and I have for some time used both the M.K.S. system and the unit pole. The usual objection to the pole concept, namely that it is a mathematical fiction, applies equally to the H, B, E and D vectors and to the scalar and vector potentials, but this is scarcely a good reason for abandoning them.

However, the method of treatment proposed is open to criticism in several respects; I shall mention three. The treatment starts with magnetostatics, but surely the stationary charge provides the logical starting-point and the electrostatic field is the simpler both conceptually and mathematically. The pole is defined in terms of permanent magnets, which are far from ideal and which introduce complicated polarization concepts. Moreover, the magnitude of the unit pole remains a mystery until eqn. (20). These difficulties are easily overcome by defining the pole in terms of a long thin coil, the ampere being defined in terms of charge.

The author includes a factor $1/\mu$ in eqn. (1); it seems to me both unnecessary and extremely unsound to base the theory of magnetic media on a definition which cannot be applied to solids. The experiment can be carried out only in fluids or gases, whose magnetic properties are of no practical interest and in which the results are ambiguous. If the pole is interpreted in terms of a coil, the validity of the $1/\mu$ factor in eqn. (1) depends on whether the fluid is admitted to, or excluded from, the inside of the coil. If we examine the electrostatic equivalent, to avoid this difficulty, we find that the force between two immersed charges has three components. The purely electric force is not affected by the medium; thus the force between the plates of a capacitor is not reduced when a dielectric is inserted. If the dielectric is a fluid and this is allowed to come into contact with the plates, it exerts a mechanical pressure* and reduces the total force to that given by eqn. (1). In general, these two force components are supplemented by a third due to the electric effect of the dipoles. It is misleading to teach that the sum of these three is comparable to the purely electric force between charges in vacuo, and it can lead to very considerable difficulties in later work.

Dr. F. T. Chapman: In Section 9 the author suggests an elementary treatment beginning with a formula for the force between magnetic poles including the factor μ, which is to be taken on trust. For many years now I have recommended that teachers, when dealing with beginnings, should give less attention to the fields of permanent magnets and should lay the major emphasis on plotting fields due to electric currents; should deal with forces between conductors and coils carrying current rather than with forces between poles; and should take as their unit the current measured by a standard balance. This provides the fourth arbitrary unit, the ampere in the M.K.S.A. system, or the deca-ampere in the C.G.S. system.

Until they have acquired sufficient knowledge of the calculus, the students will be occupied in the study of electric-circuit theory and elementary instruments and appliances. When they have the necessary mathematics they can be referred back to their plotting of the magnetic field due to, say, a circular current and have their attention drawn to the connection between the strength of the field at any point and the aspect which the circuit presents when viewed from that point. After some discussion the students can be given the definition of the magnetizing force at any point in the field due to current I in a given circuit, namely $H = Id\omega/dl$, where ω is the solid angle subtended by the circuit at this point and l is measured in the direction of H.

This definition holds for any system of units. In a rationalized system the solid angle will be measured in spheres instead of steradians and H is then in terms of ampere-turns per metre in the M.K.S.A. system. In the unrationalized C.G.S. deca-ampere system H will be in oersteds with ω measured in steradians.

Induction density is then defined from the relation $F = Bll$ or its equivalent. By simple experiments the students can now find for himself the value of $\mu = B/H$ in air, and he is then equipped with realistic ideas of the fundamental quantities he needs in order to embark upon magnetic-field theory. The language used in the subsequent development differs from that used when point poles are employed as fundamental concepts, but the process is equally rigorous.

Mr. D. J. E. Evans: A knowledge of electrical-engineering fundamentals using only the M.K.S. system of units is inadequate. Since so much historical work has been undertaken in systems of units other than the M.K.S., graduates from electrical-engineering schools should be well versed in all systems of units. When units and fundamentals are discussed it is a serious omission not to mention Professor Kapp's paper.†

Eqn. (A) is Ampère's law in the form given by Mr. Carr in the 1950 Symposium, with the exception of the introduction of a γ:

$$F = \frac{\gamma\mu_0 I_1 I_2 l}{2\pi d} \quad . \quad . \quad . \quad . \quad . \quad . \quad (A)$$

* See, for example—ABRAHAM, M., and BECKER, R.: 'Classical Theory of Electricity and Magnetism' (Blackie, 1950), p. 103.

† KAPP, R. O.: 'Differences of Opinion about Dimensions', *Proceedings I.E.E.*, Paper No. 2167, September, 1956 (104 B, p. 198).

In this equation γ is unity for a rationalized system of units and 4π for all unrationalized systems of units. Using eqn. (A) and the definition of the ampere

$$\gamma\mu_0 = 4\pi/10^7 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (B)$$

in the M.K.S. system of units, it follows that $\mu_0 = 4\pi/10^7$ if $\gamma = 1$ and $1/10^7$ if $\gamma = 4\pi$. Some may prefer to start from the unit-magnetic-pole principle. In the general representation we would say that γ units of magnetic flux are associated with unit magnetic pole. The inverse square law could then be derived as

$$F = \frac{\gamma m^2}{4\pi\mu_0 r^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (C)$$

From the definition of unit magnetic pole in the electromagnetic C.G.S. systems of units, $\gamma/\mu_0 = 4\pi$, i.e. if $\gamma = 4\pi$, $\mu_0 = 1$ (unrationalized system of e.m. units), and if $\gamma = 1$, $\mu_0 = 1/4\pi$ (rationalized system of e.m. units). Retaining the general method provides a facility of interchange between any two systems of units.

In an American paper* published in 1951 a generalized representation is used. In fact, the authors use a more general formulation, and in this it is to be preferred to the one I have described. The authors formulate Ampère's law as

$$F = \frac{n\mu_0 I_1 I_2 l}{2\pi u^2 d} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

$n = \gamma$ and is unity for all rationalized systems and 4π for all unrationalized systems. The constant u is inserted to permit both the Gaussian and Heaviside–Lorentz systems to be included in the general expression; it is equated to the velocity of light to form the relations for these two systems and to unity for all other systems.

Mr. G. F. Freeman: The engineer must ultimately think about concrete things which will work. The philosophic viewpoint propounded by the author is quite another matter, and we should bear in mind the need to evolve a theory which will not only be logical, correct and satisfying to our higher consciousness, but also capable of being put over to elementary students.

When I first received the paper I put it before several of my senior lecturers in order to get their opinions. A gentleman who is a considerable mathematician thought it perfectly delightful; the others had reservations, because they could not think of the students to whom they could put it across in its present form. So I feel that the author should have extended the Section on elementary treatment in order to make it reach a larger body of students, who, after all, are the engineers of the future.

I was very pleased that both the Measurement and Control Section and the author were brave enough to criticize some of the things which have been done in the past ten years. The M.K.S. system has its virtues, and the paper is in some ways an attempt to make an honest woman of it. But I had the good fortune to be brought up on C.G.S. units, and I have not found that they have given me mental indigestion in spite of having electrostatic units to contend with.

Kipling said 'There are five and thirty ways of reciting tribal lays, and every single one of them is right'. There are many points of view, and we want something which will satisfy our own ideas of logic. As teachers we also want something that we can put across to students. Finally, if we are engineers with no pretensions to higher mathematics, we want the plain facts. This is the range of incompatibility which we have to make compatible.

Discussions of the present type do a great deal towards clearing the air, but I should like a simplified form of the content of the paper. This is an elegant presentation, but I do not think it will reach more than a small percentage of the members of The Institution.

Finally, I would repeat what I said about rationalization some ten years ago. Rationalization means making something convenient for oneself and persuading other people that it is good for them. We are all rather too prone, particularly when we introduce something new, to let our zeal run away with our discretion and to think that what we propose must be the best. There are all sorts of ways, and those who have gone before knew what they were doing. I said that when we locked the skeleton in the cupboard it would come out again some time later. The ostriches are coming home to roost: their skeletal heads are still buried in the sand but their Achilles' heels are beginning to be visible.

Mr. T. McGreevy: Although the paper is very useful and facilitates the continuation of studies on traditional lines in which there is a very large volume of published material, I disagree completely with this method of development. The theory is built up on the twin pillars of the inverse square laws in magnetism and electrostatics. The theory has not gone very far when one is confronted with the statement that div $B = 0$ everywhere. This demolishes one pillar and the theory is left, as it were, standing on one leg. My preference is for a single foundation and I consider electrostatics as fundamental and magnetism a secondary or derived phenomenon. One of the unfortunate effects of basing theory on the magnetic pole is the tendency to imagine that permanent magnets have magnetic poles of this nature. This is quite wrong: the magnetic pole of theory is a purely mathematical concept and cannot correspond to any reality. Since magnetic effects are produced by electric currents, it is impossible to conceive of any configuration of electric charges in motion which would produce a radial magnetic field. A more practical point is that the hypothetical magnetic pole gives rise to a fixed value of magnetic flux, independent of the medium in which it is immersed; upon this basis the factor μ appears in the denominator of the inverse-square-law equation. But no one seems to have yet carried out experiments on the mutual forces between actual magnets in material media of relative permeability greater than unity—a point mentioned in Section 8. Such an experiment would not, however, support the inverse-square-law equation, since the so-called 'pole strength' would change when the medium changed. These facts seem to be a weakness in basing fundamental theory on this law of magnetism.

With regard to nomenclature and symbols, it is a pity that the author calls H the magnetic field strength; this is very confusing to students. Furthermore, the term 'permeability' is widely used to indicate the degree of penetrability of a material by a liquid or gas, and while it might have been appropriate a century ago when scientists talked about the 'electric and magnetic fluids', we now need a new word, as well as some new symbols. As the author states, μ and ϵ are mostly used for the products of a numeric and a dimensional quantity; the majority of authors use $\mu = \mu_r\mu_0$ and $\epsilon = \epsilon_r\epsilon_0$. Those who point out that μ_0 and ϵ_0 are dimensional and hence should be represented by symbols other than μ and ϵ (traditionally numerics) have tended to introduce added confusion. If the numerics are to be clearly distinguished from the quantities which have dimensions, a different *sort* of symbol should be used. The following is a suggestion which meets the valid criticisms of such advocates and yet avoids a change from the traditional usage of the symbols μ and ϵ:

$$B = \mu a H$$

$$D = \epsilon b E$$

which gives $abc^2 = 1$.

* Hessler, V. P., and Robb, D. D.: 'Generalized Electrical Formulae', *Electrical Engineering*, 1951, 70, p. 332.

Mr. L. Lewin: One of the features which a teaching of the M.K.S. system by-passes is the fact that the C.G.S. units of magnetic field and magnetic induction, the oersted and the gauss, are still used by a majority of electrical engineers and by many instrument makers. A survey of Part B of the *Proceedings*, for instance, especially papers involving such subjects as microwave isolators, will show that only a minority utilize amperes per metre in preference to oersteds. This aspect can be properly dealt with only by way of an historical development of the C.G.S. system.

The force equation between two magnetic poles can be written

$$F = \frac{m_1 m_2}{4\pi\mu r^2} = \frac{m_1 m_2}{4\pi\mu_0 r^2}\left(1 - \frac{\mu - \mu_0}{\mu}\right)$$

the two forms being algebraically equivalent, but the first exhibiting μ as a *multiplier* and the second showing its effect to be *subtractive*. Which is correct? I submit that, when the macroscopic features of a problem are under consideration, the multiplicative form is appropriate; but when we are concerned with the derivation of the formula from an examination of the microscopic aspects of the material, the subtractive form is relevant. This point can be brought out more clearly from a consideration of an analogous case, that of the commonly used formula $V = c/(\varepsilon\mu)^{1/2}$ for the velocity of light in a medium. There is no doubt that here μ and ε appear as multipliers. Nevertheless, a microscopic analysis of the properties of a loaded dielectric, for example, show that an initial wave propagates with light velocity, exciting at the scattering points wavelets whose overall effect is to cancel the initial wave and combine to form a slower one. The effect is subtractive, but we need to bother about this aspect only when the microstructure is under consideration. In the normal, macroscopic, use it is the multiplying effect which is of concern.

Mr. A. C. Sim: I should like to complain about the repeated assertion that div H 'always vanishes', and the equally disturbing statement later in the paper that 'wherever there is no charge, div $E = 0$'. While it is evident that the author is concerned with 'free space', he should state this more prominently. There is a real danger that a student reading the paper will see the emphasis on these equations but will not read the paper sufficiently carefully to note the limitation. Why speak in terms of H when physically it is B that is discussed?

It has been observed that the M.K.S. system is difficult for introduction at school. I suggest that, just as we are taught first to 'print' and then to write, so there is no reason why the obsolete systems could not be used at first and then the M.K.S. system introduced for more mature work.

Mr. K. A. Macfadyen (*communicated*): The author certainly makes progress towards the desirable goal of presenting electromagnetism in a logical way in spite of the use of M.K.S. units, but his paper serves mainly to emphasize rather than remove the central difficulty of having to leave important things undefined until a later stage in the argument is reached. More serious is the relegation of the treatment of material media to a qualitative discussion (Section 8). This evades many questions, such as the proper definition of B as $\mu_0 H + J$, where J is the magnetization of the medium. In the case of ferromagnetics eqn. (5) is no longer usable, and the corollary that $\int H \, . \, da$ vanishes in all cases is untrue. A number of magnets with their north poles inside a Gaussian surface will clearly result in a finite positive value of this integral over the surface.

Finally, it is neither necessary nor desirable to teach elementary pupils that the ratio, c, of the units of charge in the two C.G.S. systems is equal to the velocity of light (Section 9): c should be regarded as an experimentally determined constant until Maxwell's equations show its relationship with the velocity of electromagnetic waves.

Professor L. G. A. Sims (*at Portsmouth*): I agree with the author's inference that in certain ways the teaching of electromagnetism may have become more difficult by the introduction of the M.K.S. system. I have never quite believed that such physicists and mathematicians as Kelvin, Clerk Maxwell, Weber and Gauss would have agreed upon point-source electromagnetic theory unless it was a very good theory. I think also that there is really little difficulty in the concept of isolated magnetic poles simply because long thin magnets are necessary experimentally.

The merits of the M.K.S. system are that it combines the old e.m. and e.s. system and uses the magnitudes of the practical units—ohm, ampere, volt, joule, watt, henry and farad—thus eliminating one set of conversion factors. But the price paid here is that the magnitude of the velocity of light has to be introduced. These two considerations taken together mean that both permeability and permittivity acquire very odd values, both of them difficult to remember if used at longish intervals of time and both, in consequence, still little used by engineer designers, although the M.K.S. system has now been accepted for ten years. Apart from remembering the new values of constants which used to be unity, the engineer finds that algebraic expressions dealing with basic electromagnetic calculations become more cumbersome in some cases—in fact, as often as others become simplified—while basic machine-design calculations can become more liable in arithmetical error during manipulation because of the presence of small quantities less than unity.

So far as the experimental background of the M.K.S. system is concerned, those experiments which merely support the theoretical steps without conforming to high standards of accuracy are open to no objection in the context in which they are used. But it is a very different matter with experiments which are rigorous as to highest possible accuracy. These cannot proceed *pari-passu* with the M.K.S. theory. In fact, national standardizing laboratories continue to use the old classical form in realizing the units experimentally and will presumably continue to do so (unless indeed they depart radically in the direction of methods based upon atomic physics).

I suppose those of us who are power engineers accept the advantages of the M.K.S. system, but we also realize that there are practical disadvantages which make the balance between it and classical C.G.S. systems a fairly fine one. Papers which will keep discussion alive for another ten years should be welcomed at our meetings. In the meantime my own policy is to keep my students alert in how to use both systems.

Mr. Keith Morgan (*at Portsmouth*): In Southampton University, with its common course for first-year engineers, the elements of calculus can be taken for granted but quite a large percentage of the intake have no knowledge of electricity and magnetism.

It is possible to develop a common theoretical analysis which can include both C.G.S. systems and the M.K.S. unrationalized system at will. The adoption of such a system might have made the co-operation of the physicists possible, but now that rationalization is accepted we may have to contend with two separate systems. Let us avoid using units like the kilogramme-weight instead of the newton.

The paper does not explain that electrostatic fields are set up by stationary charges, and that magnetic fields and radiation fields are set up when charges move with uniform velocity and under acceleration conditions respectively.

Does the author regard B and E as the fundamental vectors of the theory as do most advanced writers? And does he think that Faraday's laws can be deduced from the rest of the theory or does he regard these as a separate premiss?

Mr. R. Yorke (*at Portsmouth*): Section 8 of the paper states that the force between magnets is inversely proportional to the permeability (of the medium), whereas that between currents is directly proportional to it. If we are to accept the electron theory of ferromagnetism, there appears to be an inconsistency here; however, it may be resolved by realizing that a current is a source of constant H, i.e. independent of the medium. Hence B, the density of the field, is proportional to the relative permeability of the medium and so also is the force on another current. Exactly the same conditions apply to the force between two permanent magnets. Between two magnetic poles as defined by eqn. (1), however, the force varies inversely as the relative permeability of the medium. These poles are sources of constant flux—independent of the medium—and so H (on which the force on this kind of pole depends) decreases as the relative permeability of the medium increases. It is the identification of the permanent-magnet pole (constant H) with this magnetic pole that is responsible for the apparent inconsistency.

Mr. O. Nourse (*at Portsmouth*): In connection with the formula for the force between conductors, I would emphasize a point that has been helpful to me, namely that problems of spherical symmetry (point sources) always involve $1/4\pi$, those of circular symmetry (parallel conductors) always involve $1/2\pi$ and those where there is an effectively parallel field involve no factor. This is the great advantage of a rationalized system.

Several speakers have firmly denied the existence of magnetic monopoles and have used this as a reason for not invoking them in what are, after all, hypothetical arguments. A recent paper* suggests two experiments for discovering the 'Dirac monopole' (the magnetic equivalent of the electron) which might occasionally occur as a 'strange' particle; the paper suggests that its value is 137/2 times the electronic charge, where 137 is the fine-structure constant.

It would seem that the formula HIl for the force on a conductor is dimensionally incorrect and should be discouraged. An experiment to prove that BIl is correct might be difficult, because the conductor must presumably have the same permeability as the medium.

The concept of the impedance of space is a very useful one and leads to a reduction in the number of constants one is required to remember; thus

$$\mu_0 = \sqrt{\frac{\mu_0}{\varepsilon_0}} \times \sqrt{(\mu_0\varepsilon_0)} = Z_0/c \simeq 377/c$$

and
$$\varepsilon_0 = \sqrt{\frac{\varepsilon_0}{\mu_0}} \times \sqrt{(\mu_0\varepsilon_0)} = \frac{1}{Z_0 c} \simeq \frac{1}{377c}$$

A further advantage is that the values of ε_0 and μ_0 cannot be interchanged accidentally.

The universal introduction of the M.K.S. system with the complete abolition of past systems would inevitably lead to the discouragement of any new systems which might be more fundamental. In this connection it is of interest to note that thinkers here and in the United States* have suggested much more fundamental systems of units 'such that the constants of space have simple relations'. Such a system has the advantage that, if the correct constants are chosen as a basis, the units so derived are no longer arbitrary. The difficulty with such systems is the usually outlandish values of length, mass and time derived which lead in some cases to even more outlandish values for electrical units. Nevertheless, one system at least, which is a semi-arbitrary system because it retains the metre and the watt,† does provide very convenient units for the communication engineer, and for the atomic physicist it has the advantage that h is nearer to unity, being numerically $5{\cdot}94 \times 10^{-17}$ ($= hc^2$ in M.K.S. units).

Mr. T. McGreevy also contributed to the discussion at Portsmouth; his remarks have been added to his contribution to the London discussion.

* Katz, R., and Parnell, D. R.: 'Two Proposed Experiments for the Detection of the Dirac Monopole', *Physical Review*, 1959, **116**, p. 236.

* Smith, P., and Nourse, O.: 'A New System of Units', *Electronic and Radio Engineer*, December, 1958, **35**, p. 480.

† Grisky, A. T.: 'Universal Units of Magnetism, Mechanics and Temperature', *Journal of the Franklyn Institute*, 1959, **268**, p. 388.

THE AUTHOR'S REPLY TO THE ABOVE DISCUSSION

Dr. P. Vigoureux (*in reply*): In the last sentence of the paper I say that I do not claim that the method presented is the only possible one, or even the only good one. Mr. Freeman's quotation says much the same thing. Accordingly, I am not surprised that a number of the speakers should prefer their own methods. I agree with Mr. Freeman's criticism of the brevity of the elementary treatment, but would point out that teachers who like the procedure indicated in that Section can extend it without difficulty.

The paper is only an introduction to the subject, very incomplete because of the need to keep it short, yet adequate in indicating the general scheme advocated. For this reason, subjects like retarded potentials have been left out; and its brevity may have justified Mr. Sim's criticism that it does not make sufficiently clear that certain formulae apply to 'free space' only. Likewise, polarization has not been treated; but on this subject Mr. Lewin has answered criticisms of several speakers as well as I could have done: it is not essential to bring in polarization when considering bulk phenomena in media which are homogeneous in bulk.

With regard to another criticism also, I am impenitent: it is better to teach beginners the M.K.S. system only. When they cease to be beginners and, having acquired enough mathematics and physics, are in a position to read works like Maxwell's or modern papers not using that system, they will have no difficulty in familiarizing themselves with the various C.G.S. systems. But so long as they are beginners they should not be burdened with tuition in those systems—an historical account free from mathematics is all that is required.

Mr. Macfadyen's Gaussian surface enclosing the north end of a magnet would have no net flux through it, as can be seen by removing an infinitely thin slice of the material where the surface cuts through the magnet: there will be a south face inside and as much flux emerges from that as from the north end of the magnet.

Mr. Yorke points out that the force between coils carrying current is proportional to the permeability of the medium in which they are immersed; this is true provided that the space inside as well as outside the coils is filled by the medium. But since with magnets it is only the space outside which is occupied by the medium, magnets must be represented by currents in solenoids from the inside of which the medium is excluded. A fairly simple argument shows that the forces are then inversely proportional to the permeability of the medium.

In reply to Mr. Morgan's queries, I have no views as to which of the vectors B or H should be regarded as fundamental, but the presentation is not affected either way. Most writers seem to consider that it is impossible to deduce the law of electromagnetic induction solely from the expression for the mutual energy of two magnetic shells or two electric circuits. The difficulty, I believe, is that when power has to be considered rather than energy, it is impossible to deduce reciprocal effects in the same rigorous manner. Induction of e.m.f. by change of flux takes place even if the circuit is not conducting; but if it is not, there is no energy involved and the derivation fails.

21.375.9

The Institution of Electrical Engineers
Paper No. 3220 E
Feb. 1960

AN INTRODUCTION TO THE THEORY OF SOLID-STATE MASERS

With Particular Reference to the Travelling-Wave Maser

By P. N. BUTCHER, Ph.D.

The paper was first received 18*th November*, 1959, *and in revised form* 2*nd January*, 1960. *It was published in February*, 1960, *and was read before the* ELECTRONICS AND COMMUNICATIONS SECTION 7*th March*, 1960.)

SUMMARY

The paper is an introduction to the theory of solid-state masers vith particular reference to the travelling-wave maser. The relevant roperties of paramagnetic ions are described and the quantum theory f maser action is outlined qualitatively. A semi-classical treatment is eveloped which is based on the classical equation of motion of a nagnetic dipole. It is used to evaluate the engineering characteristics f a travelling-wave maser which employs the comb type of slow-wave uide.

(1) INTRODUCTION

One fundamental discovery led to the development of the olid-state 'maser' (microwave amplification by stimulated emission of radiation): it was found that some solids can be mainained in an active condition with respect to a microwave requency band, i.e. they amplify radiation falling on them in his band instead of attenuating it. The solids which have this emarkable property are certain crystals doped with paramagetic ions, which are the active constituents of the crystal; ereafter we usually call them simply 'the ions', leaving undertood the fact that they are paramagnetic. The surrounding rystal holds the ions in place and modifies their behaviour in arious useful ways. The relevant properties of the ions are escribed in Section 2. The discussion is confined to ions from he iron transition group, which are the most important in ractical applications.

The ions can be activated in a variety of ways,[1] the most uccessful being the continuous microwave pumping technique uggested by Bloembergen[2] and described in Section 3, where he quantum theory of maser action is outlined qualitatively. Ve do not go into very much detail about the quantum theory, ecause the semi-classical treatment developed in Section 4 is nore illuminating. It gives one an intuitive grasp of the noneciprocal behaviour of the paramagnetic ions which is exploited n the travelling-wave maser.

Once it was appreciated that a paramagnetic crystal could be naintained in an active condition, making a practical amplifier ut of it was a relatively easy step. One simply had to put the ctive crystal into some sort of structure which carried the signal o the crystal and took it away again after it has been amplified. he obvious structure to choose is a resonant cavity with waveuide feeds, and most of the solid-state masers built to date have een of this type.[3] However, the cavity maser has several disadantages, namely bidirectional gain, instability, high gainensitivity, narrow bandwidth and tuning difficulties. It has een recognized for some time that the solution to these problems es in distributing the active crystal along a waveguide instead f putting it in a cavity, and the maser group at the Bell Telephone aboratories has recently built a travelling-wave maser with a erformance which completely fulfils the theoretical prediction.[4]

Dr. Butcher is at the Royal Radar Establishment.

The waveguide in a travelling-wave maser must have two particular properties: extremely low group velocity and distinct regions of contra-rotating circularly-polarized magnetic field. Section 5 shows how these characteristics can be achieved by using an array of parallel wires to guide the wave, taking as a specific example the comb structure used at the Bell Telephone Laboratories. Section 6 is concerned with the engineering characteristics of the travelling-wave maser, i.e. forward gain, reverse loss, bandwidth and noise figure. It is probably unnecessary to remind the reader that the current interest in masers has arisen because they have noise figures of the order of 0·1 dB.

A complete understanding of masers requires extensive use of quantum mechanics, which is not usually familiar to electrical engineers, and the semi-classical treatment which is developed here provides a useful introduction to the subject.

(2) SPIN, MAGNETIC MOMENT AND ENERGY LEVELS

The electron is not just a point having mass, m, of $0{\cdot}91 \times 10^{-27}$ g and charge, e, of $4{\cdot}8 \times 10^{-10}$ e.s.u.: it is a particle of finite size which is always spinning about an internal axis. The spin can be described by drawing a vector $\boldsymbol{s}$ pointing along the axis of rotation and having a magnitude equal to $\sqrt{3}/2$, which is the angular momentum of the electron in units of $\hbar$, Planck's constant divided by 2π (see Fig. 1). Now $\hbar = 10^{-27}$ erg-sec,

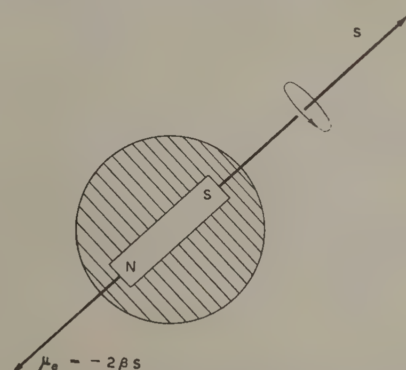


Fig. 1.—The spinning electron.

which is an extremely small angular momentum by macroscopic standards. For example, an electron moving in a circular orbit of radius 1 cm with a velocity of 10^8 cm/sec (kinetic energy 2·8 eV) has an 'orbital' angular momentum of $mvr = 0{\cdot}91 \times 10^{-19}$ erg-sec. We may therefore neglect spin when discussing the macroscopic motion of electrons. The situation is quite different when several electrons come together around a nucleus to form a stable ion. The orbit radii are then about 10^{-8} cm and the kinetic energies are of the order of 2·8 eV. Hence

$mvr = 0{\cdot}91 \times 10^{-27}$ erg-sec, which has the same order of magnitude as the spin.

In an ion it is no longer legitimate to neglect the spin in comparison with the orbital angular momentum. In fact, when the ion belongs to the iron transition group and is embedded in a crystal lattice, precisely the reverse is often true: we may neglect the orbital angular momentum in comparison with the spin. This situation occurs in the two most successful maser crystals: aluminium oxide doped with Cr^{3+} ions (ruby) and potassium cobalti-cyanide doped with Cr^{3+} ions. It comes about because the electrons re-orient their orbits in such a way that their total orbital angular momentum vanishes, in an effort to reduce their interaction with the surrounding crystal. The spins of the electrons tend to associate in anti-parallel pairs in an attempt to make the total spin vanish as well. The characteristic property of the ions of the iron transition group is that the pairing-off of the spins is incomplete. These ions have a residual total spin which can be described by a total spin vector $\boldsymbol{S}$ in the same way as we described the spin of an individual electron. The magnitude of $\boldsymbol{S}$ is equal to $\sqrt{[S(S+1)]}$, where $S = 1/2$, 1, 3/2, etc., is one half of the number of unpaired electronic spins in the ion. For an isolated electron $S = 1/2$ and $|\boldsymbol{S}| = \sqrt{3}/2$; for Cr^{3+}, $S = 3/2$ and $|\boldsymbol{S}| = \sqrt{15}/2$. It would take us too far afield to investigate the reason for this curious formula for the magnitude of the spin vector, which has its origin in the uncertainty principle. The number S is characteristic of the ion and is simply called its spin, since no confusion with the spin vector $\boldsymbol{S}$ is likely to arise. It might seem more appropriate to call $\sqrt{[S(S+1)]}$ the spin, but S is easier to remember and it is more useful in calculations.

Now that we have given up the idea that the electron is a point particle, we might try to picture it as a tiny spinning ball of charge. The spinning charge creates a magnetic field which can be calculated from Maxwell's equations. One finds that the magnetic field is that of a small bar magnet lying along the spin direction and having a magnetic moment equal to $(e/2mc)\hbar|\boldsymbol{s}|$ erg/gauss in magnitude, where $c = 3 \times 10^{10}$ cm/sec is the velocity of light (see Fig. 1). The quantity $e\hbar/2mc$ is called the 'Bohr magneton' and is denoted by $\beta = 0{\cdot}93 \times 10^{-20}$ erg/gauss. The magnet has its north pole pointing in the opposite direction to $\boldsymbol{s}$ because the electronic charge is negative (e denotes the magnitude of the electronic charge). We may describe the magnetic properties of this model of the electron by drawing a magnetic moment vector $\boldsymbol{\mu}_e = -\beta\boldsymbol{s}$, pointing in the opposite direction to $\boldsymbol{s}$ and having a magnitude equal to $\beta|\boldsymbol{s}|$. In so doing we would be wrong in only one respect: our simple model gives a magnitude to the magnetic moment which is too small by a factor of 2—the famous Landé g-factor for electron spin.[5] In fact, the magnetic moment vector of an electron is

$$\boldsymbol{\mu}_e = -2\beta\boldsymbol{s} \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

as shown in Fig. 1.

We are concerned with whole ions rather than isolated electrons. By adding together the magnetic moments of all the electrons, we obtain for the magnetic moment of an ion with spin vector $\boldsymbol{S}$,

$$\boldsymbol{\mu}_i = -2\beta\boldsymbol{S} \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

Ions of the iron transition group of elements have permanent magnetic moments because they have non-zero spin vectors. The significance of the permanent magnetic moment becomes apparent when we consider what happens when an ion is placed in a d.c. magnetic field, H_0. Suppose, for simplicity, that H_0 is directed along the z-axis of a Cartesian co-ordinate system and has magnitude H_0. Then the energy which the ion has in this field is, as for a bar magnet,[5]

$$E = -H_0\mu_{iz}$$
$$= 2\beta H_0 S_z \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

from eqn. (2). The primary distinction between a macroscopic bar magnet and an ion is that the bar magnet can take up any orientation with respect to the field, while the orientation of the ion is quantized. The quantization rule is: S_z must be integral when the spin, S, is integral and must be half-integral when S is half-integral. By taking into account the obvious geometrical restriction that $|S_z|^2 \leqslant |\boldsymbol{S}|^2 = S(S+1)$, we find that the permissible values of S_z are $-1/2$ and $+1/2$ when $S = 1/2$; -1, 0 and $+1$ when $S = 1$; $-3/2$, $-1/2$, $+1/2$ and $+3/2$ when $S = 3/2$; and $-S$, $1-S$, . . . $S-1$, S in the general case. There are $2S+1$ possible values of S_z rising from $-S$ in steps of unity. Hence the ion has a series of $2S+1$ equally spaced energy levels increasing from $-2\beta H_0 S$ in steps of $2\beta H_0$. Fig. 2 is a sketch of the possible energy levels of Cr^{3+}, for which $S = 3/2$, the levels being labelled by the values of S_z to which they correspond. With one trivial exception, this convention will be adhered to throughout the paper.

So far we have taken account only of the applied magnetic field; but there is also an internal magnetic field produced by the electrons themselves as they move in their orbits.[6] The magnetic field produced by the orbital motion of all the electrons vanishes outside the ion because both their total orbital angular momentum and their total orbital magnetic moment vanish. Inside the ion, however, the details of the orbital motion come into play and each electron experiences a finite magnetic field produced by all the others. The effect which the internal magnetic field has on the energy of the ion depends upon the angle which it makes with the total spin. Different energy levels in

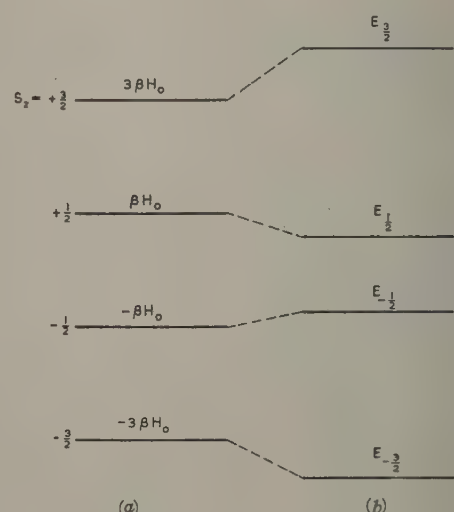


Fig. 2.—Energy levels of Cr^{3+} in a magnetic field.
(a) Neglecting internal field.
(b) Including internal field.

the external field correspond to different orientations of the tota[l] spin, and so the internal field will affect them differently. Hence one actually finds an unequally spaced series of energy levels, a[s] indicated in Fig. 2. The diagram applies to some particula[r] direction of the applied field relative to the crystal; for som[e] other direction of the applied field the energy levels will b[e]

ifferent, because there will be a change of the angles between he directions of the total spins of the various levels (which are, oughly speaking, determined by the applied field) and the irection of the internal magnetic field (which is determined by he electronic orbits and so, in the last analysis, by the surround- ng crystal). Thus, because of the internal magnetic field, the nergy levels are unequally spaced and depend upon the direction f the applied field as well as its magnitude. Unequally spaced nergy levels are essential for maser action. The anisotropy of he energy levels is a help when tuning the maser.

(3) POPULATIONS, TRANSITIONS, PUMPING AND MASER ACTION

At absolute zero the ions achieve the lowest possible total nergy by all residing in the lowest energy level. At a tempera- ure T greater than 0° K some of the ions are shaken out of the owest energy level by the thermal vibrations of the crystal. The fraction of ions in the level with $S_z = M$ is proportional to xp $(-E_M/kT)$, where $k = 1\cdot 4 \times 10^{-16}$ erg/deg is Boltzmann's onstant. Now $E_M/kT \simeq M2\beta H_0/kT = 0\cdot 16$ when $M = \frac{1}{2}$, $H_0 = 3$ kG and $T = 1\cdot 25°$ K. As a rough approximation we nay therefore expand the exponential as a power series and keep ust the first two terms. Since there are $2S + 1$ levels in all, he fraction of ions in the level with $S_z = M$ is

$$\rho_M^0 = \frac{1 - E_M/kT}{\sum_M (1 - E_M/kT)}$$

$$\simeq \frac{1 - E_M/kT}{2S + 1} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

It is interesting to digress for a moment in order to see why ons of the transition-group elements are called paramagnetic ons. In thermal equilibrium the ions prefer to occupy the ower energy levels, i.e. there are more ions with positive values f μ_{iz} than negative. Consequently, the total magnetic moment er unit volume of the crystal has a positive z-component; it is arallel to the applied field and the crystal is paramagnetic in he usual sense of the term. The paramagnetism is entirely due o the permanent magnetic moment of the transition-group ions, or ions from other groups of the Periodic Table have no per- nanent magnetic moment and Lenz's law governs their behaviour n an applied magnetic field. The electronic orbits precess so s to set up a magnetic moment opposing the applied field, i.e. he ions are diamagnetic.

Paramagnetic ions can be induced to jump from one level to nother by subjecting them to a magnetic field oscillating at an ngular frequency equal to the energy difference divided by $\hbar$. For adjacent levels the transition frequency is (neglecting the nternal magnetic field) $2\beta H_0/\hbar = 2\pi \times 2\cdot 8 H_0$ Mrad/sec, which s a microwave frequency when H_0 is a few kilogauss. It would ppear from what has been said that the ions respond to a nicrowave magnetic field only at precisely the transition fre- quency, but this is not true, because there is yet another internal nagnetic field acting on the ions—the field which the ions roduce outside themselves by virtue of their spin magnetic noments.[6] Each ion experiences the magnetic field produced y all the other ions and any nuclear magnetic moments which nay be in the crystal. This field is different at the sites of lifferent ions and so the energy levels are spread out over a and. The ions therefore have a bell-shaped frequency-response urve, the precise shape of which is determined by the density of evels in the various parts of the energy bands. It is called the ine shape and is about 50 Mc/s wide in good maser crystals.

When a particular transition is excited by microwave radiation with angular frequency ω, ions in the lower level jump up to the upper level by absorbing a quantum $\hbar\omega$ of radiation energy, and ions in the upper level fall to the lower level by emitting a quantum of radiation energy. The transition rate per ion, V, is the same in both directions.[1] Consequently the net power absorbed from the radiation per unit volume of crystal is

$$P_m = \hbar\omega N(\rho_l - \rho_u)V \quad . \quad . \quad . \quad . \quad . \quad (5)$$

where N is the number of ions per unit volume, ρ_l is the fraction of ions in the lower level and ρ_u is the fraction of ions in the upper level. The important thing to notice is that the absorbed power is proportional to the population difference across the transition $N(\rho_l - \rho_u)$.

When the ions are in thermal equilibrium, $\rho_l > \rho_u$ and there is a net positive absorption of radiation, i.e. the crystal is passive. In a maser the thermal-equilibrium populations are upset in such a way that $\rho_l < \rho_u$. The incident radiation then stimulates a net negative absorption, i.e. a net emission of extra radiation, and the crystal is active. There are several ways of achieving this condition,[1] the most satisfactory method being that suggested be Bloembergen[2] which utilizes three energy levels of an ion with a spin of one or more. Let us number these levels 1, 2 and 3 in order of increasing energy (see Fig. 3). The signal to be

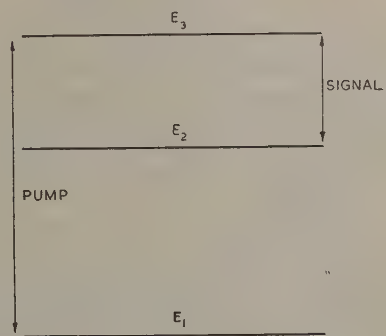


Fig. 3.—The 3-level maser.

amplified induces transitions between levels 2 and 3. The population of level 3 is made greater than that of level 2 by 'pumping' ions up from level 1 with an intense microwave field at the resonant frequency of the 1–3 transition.

The pumping action comes about as follows. When the pump field has a low intensity, the populations are not much disturbed from their thermal-equilibrium values; the ions there- fore absorb power from the pump, because more ions jump from level 1 to level 3 than fall from level 3 to level 1. Obviously, the population of level 3 is increased by the absorption process, but it is prevented from increasing indefinitely by the thermal vibrations of the crystal, which do their best to restore thermal equilibrium. The net result of the competition between the pump and the thermal vibrations is that the populations reach steady values with rather more ions in level 3 and rather less ions in level 1 than is the case in thermal equilibrium. As the pump intensity is raised, the population of level 3 increases and the population of level 1 decreases until, when the pump completely overcomes the thermal vibrations, both levels have the same populations.

Efficient pumping of suitable crystals makes the population of level 3 greater than that of level 2 and the ions become active at the signal frequency. Indeed, when levels 1 and 3 are equally populated, the population differences across the 2–3 transition

and the 1–2 transition are equal in magnitude and opposite in sign. Hence the crystal is active either at the frequency of the 2–3 transition, as assumed in Fig. 3, or at the frequency of the 1–2 transition. Efficient pumping activates the crystal at one frequency or the other (disregarding the unlikely event of equal populations in all three levels). The 3-level maser scheme fails, in principle, only if the energy levels are equally spaced, when the signal stimulates both the active and the passive transitions. The net effect is a small absorption or emission depending on which transition has the largest transition rate. Thus the unequally spaced energy levels produced by the internal magnetic field are essential for maser action.

We can easily estimate the fractional population difference across the 2–3 transition when the pump field is intense. Let ρ_1^0, ρ_2^0 and ρ_3^0 be the fractional populations of the three levels in thermal equilibrium. If the pump affects only the populations of levels 1 and 3, then the populations after pumping are

$$\rho_3 = \rho_1 = \tfrac{1}{2}(\rho_1^0 + \rho_3^0)$$
$$\rho_2 = \rho_2^0$$

Hence, by using eqn. (4) with a slight change of notation, we obtain

$$\rho_2 - \rho_3 = \rho_2^0 - \tfrac{1}{2}(\rho_1^0 + \rho_3^0)$$
$$= \frac{\tfrac{1}{2}(E_1 + E_3) - E_2}{(2S+1)kT} \quad . \quad . \quad . \quad . \quad (6)$$

which is negative provided that level 2 is nearer to level 3 than it is to level 1. The numerator in this expression is typically of the order of $-\tfrac{1}{2}\beta H_0$, so that $\rho_2 - \rho_3 \simeq -\tfrac{1}{2}\beta H_0/[(2S+1)kT] = -2{\cdot}0\%$ when $H_0 = 3\,\mathrm{kG}$, $S = 3/2$ and $T = 1{\cdot}25°\,\mathrm{K}$. Clearly, the population difference after pumping has the same order of magnitude as the thermal-equilibrium population difference, but its sign has been changed by the pump.

The power necessary for efficient pumping is determined by the strength of the coupling between the thermal vibrations and the ions, the amplitude of the thermal vibrations and the density of ions. Some ions are more weakly coupled to the thermal vibrations than others. Cr^{3+} in ruby or potassium cobalticyanide is particularly good in this respect. The amplitude of the thermal vibrations can be reduced by cooling the crystal, usually to liquid-helium temperatures. Maser action has been observed in ruby at 60° K, but the effect is less pronounced because the temperature appears in the denominator of expression (6) for the fractional population difference across the active transition.[7] The density of ions also plays a significant role in determining the pump power: as the density is raised, the ions come closer together and couple more strongly through their magnetic fields, and it then becomes more difficult to disturb the thermal equilibrium of the ions. For example, dark ruby containing 1% Cr^{3+} or more is unaffected at the pump power level required to pump effectively light ruby containing about 0·05% Cr^{3+}. A similar situation exists in other paramagnetic crystals. We see that pumped paramagnetic crystal comes in two grades, namely active crystal (low density) and passive crystal (high density); this fact is exploited in the travelling-wave maser.

(4) SEMI-CLASSICAL THEORY OF MASER ACTION

The quantum theory of maser action has been outlined in Section 3. To complete the analysis we have to calculate the transition rate induced by the signal field. While it is not difficult to do this quantum-mechanically,[6, 8] it is more illuminating to carry out a classical calculation of the response of the ions to the signal field. The classical approach has one defect: we cannot easily take into account the internal magnetic field whic is, in fact, essential for maser action. The difficulty can be ove come by a trick which gives precisely the right answer in son instances and gives the right physical picture in the general cas The classical theory is more illuminating than the standar quantum-mechanical calculation; it gives one an intuitive gras of the non-reciprocal behaviour of the paramagnetic ions whic is exploited in the travelling-wave maser.

First we give a classical description of the behaviour of a ion with spin vector S when it is subjected to a d.c. magnet field H_0. This topic has already been covered from a quantum mechanical point of view, but a classical treatment is necessar before we attempt to handle the response to the signal field b classical methods. The angular momentum of the ion in c.g. units is $\hbar S$, and the magnetic moment is $\mu_i = -2\beta S$. The d. magnetic field exerts a couple on the magnetic moment which at right angles to both H_0 and μ_i and has magnitude $H_0|\mu_i| \sin \theta$ where θ is the angle between H_0 and μ_i. In vector notation th couple is $\mu_i \times H_0$, the vector product of μ_i and H_0. Newton' law for rotational motion states that the rate of change of th angular momentum is equal to the couple. Hence

$$\hbar\frac{d}{dt}S = \mu_i \times H_0 \quad . \quad . \quad . \quad . \quad . \quad (7$$

i.e. since $S = -\mu_i/2\beta$,

$$\frac{d}{dt}\mu_i = \gamma H_0 \times \mu_i \quad . \quad . \quad . \quad . \quad . \quad (8$$

where $\gamma = 2\beta/\hbar = 2\pi \times 2{\cdot}8$ Mrad/gauss is the 'gyro-magneti ratio' of the ion.

The solution of eqn. (8) is well known, and is illustrated i Fig. 4. The magnetic moment precesses about the d.c. magneti

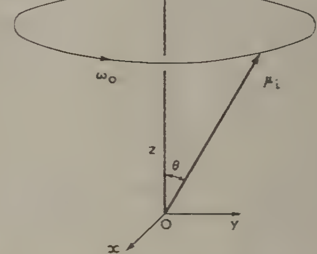


Fig. 4.—The precessing magnetic moment.

field with an angular velocity $\omega_0 = \gamma H_0$. The sense of rotatio is related to the direction of H_0 in the same way as the sense o rotation and the direction of motion when driving a right handed screw; for brevity, we call this a positive rotation abou H_0, and that in the opposite sense is called a negative rotation The angle θ, between μ_i and H_0 can have any value. The validity of this solution can be verified by substitution in eqn. (8)

We have just seen that the angle between μ_i and H_0 is arbitrary This conclusion appears to contradict completely the quantization rule for the orientation of the magnetic moment; nevertheless, the classical approach is rigorously correct in this respect We saw in the quantum-mechanical treatment that the ions ar not all in the same state, some having one allowed orientatio and some another. The uncertainty is removed when we consider (as we must do) a large number of ions. Their mea magnetic moment has a well-defined direction, but since it is a mean, it need not have any of the allowed orientations and ca

make any angle with H_0. The mean magnetic moment is the quantum-mechanical analogue of the classical magnetic moment which appears in eqn. (8).

Eqn. (8) is not the whole story; obviously there is something missing, because it predicts that the magnetic moment will go on precessing about H_0 for ever, whereas we know that, in fact, it will lower its energy so far as possible by coming into line with H_0. The coupling of the ion to the other ions and to the surrounding crystal causes this relaxation of the precessing moment towards an equilibrium value. A description of the relaxation process which is adequate for our purpose can be obtained by adding to the right-hand side of eqn. (8) the extra term $-(\boldsymbol{\mu}_i - \boldsymbol{\mu}_0)/T_2$, where $\boldsymbol{\mu}_0$ is the equilibrium moment which is parallel to H_0 and T_2 is the relaxation time (strictly the spin-spin relaxation time).[9] The equation then becomes

$$\frac{d}{dt}\boldsymbol{\mu}_i = \gamma \boldsymbol{H}_0 \times \boldsymbol{\mu}_i - (\boldsymbol{\mu}_i - \boldsymbol{\mu}_0)/T_2 \quad . \quad . \quad . \quad (9)$$

Since $\boldsymbol{\mu}_0$ is parallel to H_0, eqn. (9) obviously has the steady-state solution $\boldsymbol{\mu}_i = \boldsymbol{\mu}_0$. If the magnetic moment is moved away from the equilibrium position, it spirals back towards it, rotating at the precession frequency ω_0 and decreasing the angle of precession exponentially with a time-constant equal to T_2. To prove that this is so, we have only to verify by substitution in eqn. (9) that it has the general solution

$$\boldsymbol{\mu}_i = \boldsymbol{\mu}_0 + \exp(-t/T_2)\boldsymbol{\mu}'$$

where $\boldsymbol{\mu}'$ is a solution of eqn. (8).

We are now ready to calculate the response of the ion to the magnetic field of the microwave signal

$$H_1 = \mathscr{R}(H\varepsilon^{j\omega t}) \quad . \quad . \quad . \quad . \quad . \quad (10)$$

where $H = (H_x, H_y, H_z)$ is the complex vector amplitude of the field and $\mathscr{R}$ denotes 'the real part of'. The equation of motion of $\boldsymbol{\mu}_i$ in the presence of H_1 is obtained from eqn. (9) by substituting the total field $H_0 + H_1$ for H_0 alone. When H_1 is small, $\boldsymbol{\mu}_i$ is perturbed only slightly from $\boldsymbol{\mu}_0$ and becomes $\boldsymbol{\mu}_0 + \boldsymbol{\mu}_1$. By multiplying out the vector product and neglecting the second-order term $H_1 \times \boldsymbol{\mu}_1$, we obtain the equation of motion of the perturbation $\boldsymbol{\mu}_1$ as

$$\begin{aligned} d\boldsymbol{\mu}_1/dt &= \gamma(H_0 \times \boldsymbol{\mu}_1 + H_1 \times \boldsymbol{\mu}_0) - \boldsymbol{\mu}_1/T_2 \\ &= \gamma H_0 \times (\boldsymbol{\mu}_1 - KH_1) - \boldsymbol{\mu}_1/T_2 \quad . \quad . \quad (11) \end{aligned}$$

since $\boldsymbol{\mu}_0 = KH_0$, where K is some proportionality constant. The steady-state solution of eqn. (11) may be written in the form

$$\boldsymbol{\mu}_1 = \mathscr{R}(\boldsymbol{\mu}\varepsilon^{j\omega t}) \quad . \quad . \quad . \quad . \quad . \quad (12)$$

where $\boldsymbol{\mu} = (\mu_x, \mu_y, \mu_z)$ is the complex vector amplitude of $\boldsymbol{\mu}_1$. By taking the z-axis along H_0 (see Fig. 4) and writing out the Cartesian components of eqn. (11) with $\boldsymbol{\mu}_1$ expressed as in eqn. (12) we obtain

$$\left.\begin{aligned} \left(j\omega + \frac{1}{T_2}\right)\mu_x &= -\omega_0(\mu_y - KH_y) \\ \left(j\omega + \frac{1}{T_2}\right)\mu_y &= +\omega_0(\mu_x - KH_x) \\ \left(j\omega + \frac{1}{T_2}\right)\mu_z &= 0 \end{aligned}\right\} \quad . \quad . \quad (13)$$

Hence $\mu_z = 0$, while μ_x and μ_y are determined by the first pair of equations in the set (13). Instead of solving them directly, it is more instructive to work with the quantities

$$\left.\begin{aligned} \mu^+ &= \mu_x + j\mu_y \\ \mu^- &= \mu_x - j\mu_y \\ H^+ &= H_x + jH_y \\ H^- &= H_x - jH_y \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad (14)$$

These quantities have a simple interpretation to be explained later. By multiplying the second equation in the set (13) by j and then adding and subtracting it from the first equation, we obtain immediately explicit formulae for μ^+ and μ^-, namely

$$\mu^+ = \frac{-jK\omega_0 H^+}{j(\omega - \omega_0) + 1/T_2} \quad . \quad . \quad . \quad (15)$$

$$\mu^- = \frac{+jK\omega_0 H^-}{j(\omega + \omega_0) + 1/T_2} \quad . \quad . \quad . \quad (16)$$

Before going on we must see what the quantities introduced by eqn. (14) actually represent. The first thing to notice is that μ_z and H_z do not appear in eqns. (15) and (16); μ_z is zero and H_z does not couple to the magnetic moment. In what follows we shall ignore H_z, so that $\boldsymbol{\mu}_1$ and H_1 both lie in the xy-plane. It is sufficient to consider H_1 alone. By using the definitions (10) and (14) and ignoring H_z we have

$$\begin{aligned} H_1 &= \mathscr{R}[(H_x, H_y)\varepsilon^{j\omega t}] \\ &= \mathscr{R}\left\{\left[\frac{H^+ + H^-}{2}, \frac{-j(H^+ - H^-)}{2}\right]\varepsilon^{j\omega t}\right\} \\ &= \mathscr{R}[\tfrac{1}{2}H^+(1, -j)\varepsilon^{j\omega t}] + \mathscr{R}[\tfrac{1}{2}H^-(1, j)\varepsilon^{j\omega t}] \quad . \quad (17) \end{aligned}$$

In the last line H_1 has been broken up into the superposition of two fields; the first is

$$\begin{aligned} H_1^+ &= \mathscr{R}[\tfrac{1}{2}H^+(1, -j)\varepsilon^{j\omega t}] \\ &= \tfrac{1}{2}|H^+|(\cos\phi, \sin\phi) \quad . \quad . \quad . \quad (18) \end{aligned}$$

where $\phi = \omega t$ plus the phase angle of H^+ and increases as time progresses. When $\phi = 0$, H_1^+ is directed along the x-axis; a quarter-cycle later, when $\phi = \frac{1}{2}\pi$, H_1^+ is directed along the y-axis (see Fig. 4). Moreover, H_1^+ obviously has the constant magnitude $|H_1^+| = \frac{1}{2}|H^+|$. We conclude that H_1^+ is a circularly polarized field rotating in the positive sense about 0z. The second part of H_1, namely

$$H_1^- = \mathscr{R}[\tfrac{1}{2}H^-(1, j)\varepsilon^{j\omega t}] \quad . \quad . \quad . \quad . \quad (19)$$

is also circularly polarized, but rotates in the negative sense because of the different sign in front of j.

We see that H^+ and H^- specify the amplitude and phase of the two contra-rotating circularly polarized fields into which the microwave magnetic field can be analysed. Similarly, μ^+ and μ^- specify the amplitude and phase of the two contra-rotating circularly polarized components ($\boldsymbol{\mu}_1^+$ and $\boldsymbol{\mu}_1^-$) which together make up the perturbation of the magnetic moment. The resolution of H_1 and $\boldsymbol{\mu}_1$ into circularly polarized components enters naturally into the solution of the equation of motion, because the natural motion of the magnetic moment is, apart from relaxation, circularly polarized. Moreover, we know that the natural motion is a rotation in the positive sense at the precession frequency ω_0. Consequently we expect to find that $\boldsymbol{\mu}_1^+$ is strongly excited by H_1^+ when $\omega \simeq \omega_0$, while $\boldsymbol{\mu}_1^-$ is only weakly excited by H_1^- at all frequencies. These expectations are realized in eqns. (15) and (16): μ^+ is excited by H^+ and exhibits a resonance when $\omega = \omega_0$, while μ^- is excited by H^- and does not exhibit a resonance.

To summarize the results of the above discussion we can say that when $\omega \simeq \omega_0$ the ions couple only to the circularly polarized

component of the magnetic field which rotates in sympathy with the natural precessional motion, and the coupling is a maximum when the frequency of the field is equal to the frequency of precession. It follows from this fact that a paramagnetic crystal is a non-reciprocal transmission medium for circularly polarized electromagnetic waves. Consider what happens when a wave rotating in the positive sense about the direction of propagation passes through the crystal, first in the direction of H_0 and then in the opposite direction. In the first case the microwave magnetic field rotates in sympathy with the precessional motion, couples strongly to the ions and exchanges energy with them. In the second case the microwave magnetic field rotates in the opposite sense to the precessional motion, does not couple to the ions and does not exchange energy with them. This is what we mean by a non-reciprocal transmission medium. We may contrast the behaviour of a paramagnetic crystal with that of lossy dielectric, which is a typical reciprocal medium. The attenuation of the wave by the lossy dielectric is the same for both directions of propagation, because it depends only on the magnitude of the electric field and is independent of the sense of rotation. The non-reciprocal behaviour of the paramagnetic ions is used in the travelling-wave maser to obtain gain in one direction and attenuation in the opposite direction.

We are principally interested in the power absorbed from the microwave field by the magnetic moment. The absorbed power is given by

$$p_m = \mathrm{Av}\,(\boldsymbol{H}_1 \,.\, d\boldsymbol{\mu}_1/dt) \quad . \quad . \quad . \quad . \quad (20)$$

where Av denotes 'time-average' and the dot denotes the product of the vectors' magnitudes and the cosine of the angle between them. Eqn. (20) is the magnetic analogue of the mean potential difference times current formula for electrical circuits. To evaluate p_m we write $\boldsymbol{H}_1 = \boldsymbol{H}_1^+ + \boldsymbol{H}_1^-$ as in eqn. (17) and use the corresponding resolution of $\boldsymbol{\mu}_1 : \boldsymbol{\mu}_1^+ + \boldsymbol{\mu}_1^-$. Thus we obtain

$$p_m = \mathrm{Av}\,[(\boldsymbol{H}_1^+ + \boldsymbol{H}_1^-)\,.\,d(\boldsymbol{\mu}_1^+ + \boldsymbol{\mu}_1^-)/dt]$$

We can neglect both $\boldsymbol{\mu}_1^-$ and $\boldsymbol{H}_1^-$ in this expression, because $\boldsymbol{\mu}_1^-$ is negligible compared with $\boldsymbol{\mu}_1^+$ near resonance and $\boldsymbol{H}_1^-$ rotates in the opposite sense to $d\boldsymbol{\mu}_1^+/dt$, so that their scalar product averages to zero. Hence

$$p_m = \mathrm{Av}\,(\boldsymbol{H}_1^+ \,.\, d\boldsymbol{\mu}_1^+/dt) \quad . \quad . \quad . \quad . \quad (21)$$

Now both $\boldsymbol{H}_1^+$ and $d\boldsymbol{\mu}_1^+/dt$ rotate in the same direction with the same angular velocity, keeping the same lengths and the same angle between them; consequently, $\boldsymbol{H}_1^+ \,.\, d\boldsymbol{\mu}_1^+/dt$ is, in fact, constant. The lengths of the two vectors are $\frac{1}{2}|H^+|$ and $\frac{1}{2}\omega|\mu^+|$ and the angle between them is $\psi = \frac{1}{2}\pi$ + phase angle of μ^+ − phase angle of H^+; the ω and $\frac{1}{2}\pi$ appear because of the time derivative. Hence, by using the rules for complex multiplication and substituting for μ^+ from eqn. (15), we obtain

$$\begin{aligned} p_m &= \frac{\omega}{4}|H^+|\,|\mu^+|\cos\psi \\ &= \mathscr{R}(j\omega\overline{H^+}\boldsymbol{\mu}^+/4) \\ &= \frac{K\omega\omega_0}{4}|H^+|^2\frac{1/T_2}{(1/T_2)^2 + (\omega-\omega_0)^2} \quad . \quad (22) \end{aligned}$$

where $\overline{H^+}$ denotes the complex conjugate of H^+. Eqn. (22) is essentially the desired formula, but it is convenient to make a few changes at this point to put the formula into a more useful form. First, we multiply by N, the number of ions per unit volume, to obtain the power absorbed per unit volume. Then we multiply and divide by $2T_2^2$ to put the last factor into its standard form. Finally, we write $\omega_0 = \gamma H_0 = 2\beta H_0/\hbar$ ar $KH_0 = \mu_0$, the z-component of the equilibrium moment. Tl final result for the power absorbed by the paramagnetic io per unit volume of the crystal is

$$P_m = Np_m = \hbar\omega(N\mu_0\beta/4\hbar^2)|H^+|^2\frac{2T_2}{1+T_2^2(\omega-\omega_0)^2} \quad (2$$

The last factor in the expression for P_m specifies the frequen dependence of the absorption, i.e. the line shape. The wid of the line between the half-power points is $2/T_2$ rad/sec = $1/(\pi T_2)$ c/s, which is about 50 Mc/s in active maser crystal while $\omega/2\pi \simeq \omega_0/2\pi$ is a microwave frequency. It is therefo legitimate to neglect the frequency dependence of the ω in tl first factor of P_m when considering the line shape near resonanc

At the beginning of this Section it was remarked that it wou be necessary to use a trick to get the right answer, because v are neglecting the internal magnetic field. There has been r trickery so far; it becomes necessary when we evaluate $N\mu_0$, tl z-component of the equilibrium magnetic moment per ur volume—which is the only unknown quantity in the formula fo P_m. The difficulty arises because the energy levels are equal spaced in the absence of the internal magnetic field. Cons quently, the signal field stimulates transitions between every pa of adjacent levels, whereas in real maser crystals the levels ar unequally spaced and only one transition is excited. Th unrealistic feature of our model can be obviated by ensuring th only the transition we are interested in absorbs power, in spi of the fact that the levels are equally spaced.

To see how this can be achieved we consider the case of tl transition between the $-\frac{1}{2}$ and $+\frac{1}{2}$ levels in Cr^{3+} (see Fig. 5 The quantum-mechanical theory in Section 3 showed that tl net power absorbed by a transition is proportional to tl population difference across it [see eqn. (5)], being positive whe there are more ions in the lower level than there are in tl upper level, negative when the reverse is true and zero when tl two populations are equal. With this in mind, we consider a artificial 'equilibrium' situation in which the fractional popul tions of the various levels have the values shown in Fig. 5. Tl

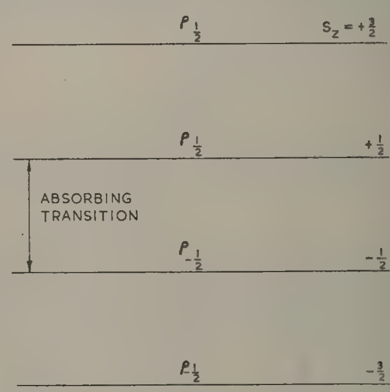


Fig. 5.—Artificial populations which suppress the $-\frac{3}{2}$ to $-\frac{1}{2}$ and $+\frac{1}{2}$ to $+\frac{3}{2}$ transitions of Cr^{3+}.

fractional populations $\rho_{-1/2}$ and $\rho_{1/2}$ are determined by requiri that $\rho_{-1/2} - \rho_{1/2}$ has the value actually established by the pun and that $2(\rho_{-1/2} + \rho_{1/2}) = 1$, i.e. the sum of the fraction populations is unity as it should be. In this situation we ha the true population difference across the $-1/2$ to $+1/2$ tran tion and zero population difference across the other two tra sitions stimulated by the signal field. Consequently, all t

power absorbed by the ions goes into the $-\frac{1}{2}$ to $+\frac{1}{2}$ transition, and the power absorbed by this transition alone is obtained by substituting in eqn. (23) the value of $N\mu_0$ appropriate to the artificial equilibrium situation.

From Fig. 5 we see that, of the N ions per unit volume, a fraction $\rho_{1/2}$ have $S_z = \frac{3}{2}$, a fraction $\rho_{1/2}$ have $S_z = \frac{1}{2}$, a fraction $\rho_{-1/2}$ have $S_z = -\frac{1}{2}$ and a fraction $\rho_{-1/2}$ have $S_z = -\frac{3}{2}$. Hence, from eqn. (2), the z-component of the 'equilibrium' moment per unit volume is

$$N\mu_0 = -2\beta N(\tfrac{3}{2}\rho_{1/2} + \tfrac{1}{2}\rho_{1/2} - \tfrac{1}{2}\rho_{-1/2} - \tfrac{3}{2}\rho_{-1/2})$$
$$= 4\beta N(\rho_{-1/2} - \rho_{1/2}) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (24)$$

By substituting this value of $N\mu_0$ in eqn. (23) we obtain the final result for the power absorbed by the $-\frac{1}{2}$ to $+\frac{1}{2}$ transition as

$$P_m = \hbar\omega . N(\rho_{-1/2} - \rho_{1/2})\left[(\beta/\hbar)^2|H^+|^2\frac{2T_2}{1+T_2^2(\omega-\omega_0)^2}\right] \quad (25)$$

The expression has been broken up into the product of the three factors which occur in eqn. (5), namely the microwave quantum, the population difference per unit volume and the transition rate per ion. A quantum-mechanical calculation yields precisely the same expression, provided that it is correct to label the energy levels with values of S_z as we have done. Actually this is true only in special cases (e.g. H_0 along the c-axis in ruby) because the internal magnetic field usually mixes up the S_z values, but that is a technical complication of secondary importance from our point of view.

The same technique can be applied in the general case to calculate the power absorbed by the transition from $S_z = M$ to $S_z = M + 1$ when the ions have spin S. The result is

$$P_m = \hbar\omega N(\rho_M - \rho_{M+1}) \times$$
$$\left[(\beta/2\hbar)^2(S-M)(S+M+1)|H^+|^2\frac{2T_2}{1+T_2^2(\omega-\omega_0)^2}\right] \quad (26)$$

which reduces to eqn. (25) when $M = -\frac{1}{2}$ and $S = \frac{3}{2}$. The author hopes that the experts will be as delighted as he was to find the matrix element of the magnetic moment appearing correctly in this expression.

(5) THE COMB STRUCTURE

We have seen how to obtain the active crystal: to make an amplifier from it, we have simply to put the crystal in some sort of structure to carry the signal power (and the pump power) to the active crystal and take it away again after it has been amplified. The obvious structure to choose is a resonant cavity with waveguide feeds (see Fig. 6); this has the advantage of producing large fields in the cavity for small input powers, but it has a number of disadvantages, notably

(*a*) Small gains are easily obtained, but to achieve the 20–30 dB gain required of a practical device, the coupling to the waveguide feeds must be reduced so that less power escapes from the cavity and the field strength in it builds up.[10] If the coupling is reduced too far, the maser breaks into oscillation because the emission from the active crystal stimulated by noise in the cavity is more than enough to supply the ohmic losses and the waveguide feeds. High gain is obtained by operating the maser on the verge of oscillation, which is obviously a most sensitive and unstable condition.

(*b*) As the gain increases the amplification bandwidth narrows, falling off as the inverse square root of the gain.[10] At 20 dB gain the maximum possible bandwidth of a cavity maser (reflection type) is 20% of the line width, i.e. typically 10 Mc/s.

(*c*) It is a simple matter to tune the transition frequencies of the paramagnetic ions by altering the magnitude and direction of the d.c. magnetic field, but the cavity must be tuned at the same time—which is not so simple.

(*d*) The amplification is reciprocal, i.e. power sent through the cavity in either direction is amplified by the same factor.[10] Consequently, small reflections in the input and output waveguides of a high-gain maser can easily send it into oscillation. The reciprocity of the gain may be surprising in view of the fact that the paramagnetic crystal is non-reciprocal; it comes about because the magnetic field in a cavity is linearly polarized, i.e. it consists of two contra-rotating circularly polarized components of equal magnitude. Whatever direction the power is coming from, one of these circularly polarized components couples strongly to the paramagnetic ions. Non-reciprocal gain is achieved in cavity masers by adding non-reciprocal ferrite elements outside the cavity.

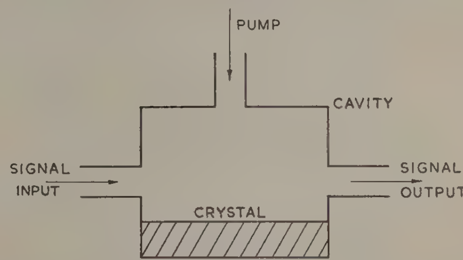


Fig. 6.—The transmission cavity maser.

All the above disadvantages of putting the active crystal in a cavity are alleviated to a considerable extent when the active crystal is disturbed along a waveguide instead (see Fig. 7). Thus,

(*e*) High gain can be achieved by increasing the length of waveguide and by decreasing the velocity at which energy propagates through it, so as to keep the signal in the active crystal for a long time. Neither procedure leads to excessive gain sensitivity or instability.

(*f*) High gain results from an exponential growth along the waveguide instead of from positive feedback. Consequently, as we shall see later, the bandwidth falls off as the inverse square root of the gain in decibels instead of falling off as the inverse square root of the actual gain.

(*g*) Tuning the paramagnetic ions alone will tune the entire device over the pass-band of the waveguide.

(*h*) The amplification can be made non-reciprocal by placing the active crystal in regions of circularly polarized magnetic field. In fact, as we shall see later, by putting active crystal (low density) and passive crystal (high density) in distinct regions of contra-rotating circular polarization we can obtain gain in one direction and an equal or larger loss in the opposite direction. The device cannot oscillate under these conditions, however large the reflections in the input and output waveguides may be.

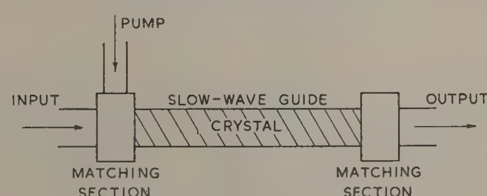


Fig. 7.—The travelling-wave maser.

The essential characteristics desired of the waveguide structure for a travelling-wave maser are low velocity of energy propagation (group velocity) and distinct regions of contra-rotating circular polarization, and both are possessed by structures in which the wave is guided along an array of parallel wires set transverse to the direction of propagation. Structures of this type have already found many applications to electronic travelling-wave tubes in which the active element is an electron beam, e.g. helices, ladder lines, meander lines, interdigital lines and combs.[11] We shall consider the comb structure because it has been used successfully in a travelling-wave maser at Bell Telephone Laboratories.[4]

The 'pure' comb consists of an array of identical, equally spaced, parallel wire fingers, short-circuited at one end and open-circuited at the other (see Fig. 8). However, the pure comb will not transmit electromagnetic energy along its length. If we excite a particular finger, the wave bounces backwards and forwards between the short-circuit at one end and the open-circuit at the other; it is not passed on to adjacent fingers. In other words, the group velocity is zero for a pure comb. For our purpose this is an advantage, because, by introducing a small perturbation, such as the capacitive loading plate indicated in Fig. 8, we can provide a connecting link between the fingers and

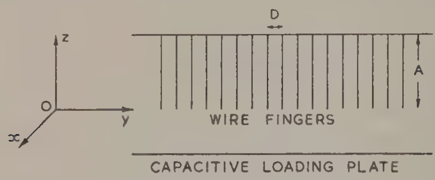


Fig. 8.—The comb structure.

propagate a signal along the comb (in the y-direction) with a very small group velocity. We shall see that group velocities of about 1% of the velocity of light are required. This represents a very high degree of wave-slowing, which the comb structure is ideally suited to perform.

The finger-length, A, appropriate to the signal frequency ω is determined by the following considerations. When the pure comb is excited at the signal frequency a standing-wave is set up on each finger consisting of two waves travelling along it in opposite directions with the velocity of light. Since there is an open-circuit at the finger tip, the current along it varies as $\sin(2\pi z/\lambda)$, where z is measured from the finger tip and λ is the free-space wavelength at the signal frequency. At the base of the finger there is a short-circuit, i.e. a current maximum. Hence $2\pi A/\lambda = \pi/2$ and $A = \lambda/4$, i.e. the fingers must be a quarter-wavelength long if the pure comb is to be capable of supporting a field at the signal frequency.

We see that the pure comb has an infinitesimal pass-band, limited to the frequency for which the finger length is precisely $\lambda/4$. The capacitive loading plate produces a finite pass-band. We may estimate its width very roughly by remembering that the group velocity is defined by the equation[12]

$$v_g = d\omega/d\beta_g \quad . \quad . \quad . \quad . \quad . \quad . \quad (27)$$

where β_g is the phase-change coefficient in the direction of propagation. Now the comb is a periodic structure; the pass-band limits are therefore $\beta_g = 0$ and $\beta_g = \pi/D$, where D is the period of the comb.[12] Very roughly, then, the width of the pass-band is

$$\frac{\Delta\omega}{2\pi} = \frac{d\omega}{d\beta_g}\frac{\Delta\beta_g}{2\pi}$$

$$= \frac{v_g \pi/D}{2\pi} \quad . \quad . \quad . \quad . \quad . \quad . \quad (28)$$

which is 750 Mc/s when $v_g = c/100$ and $D = 2$ mm. The pass-band is very much wider than the line width of the paramagnetic crystal, and it provides a large tuning band for the travelling-wave maser which can be increased by reducing the period of the comb.

The magnetic field distribution is hardly disturbed by the capacitive loading plate, but even so, the calculation of the details of the field distribution is obviously a formidable task.[11] Before an approximate solution of the problem is derived, it is useful to observe one or two simple properties of the fie The magnetic field encircles the fingers and the electric fie points radially from them; hence both fields are in the xy-pla (see Fig. 8). Moreover, the x-component of the magnetic fie is symmetrical about the plane of the comb while the y-co ponent is anti-symmetrical. The last two statements a obviously true close to the fingers, but they apply quite genera because, apart from a small end-effect due to the capaciti loading plate, there is no geometrical feature to render the invalid as we move away from the comb. The symmet properties allow us to confine our attention to the field bel the comb ($x < 0$).

The essential properties of the magnetic field distributi below the comb can be obtained by recognizing that the prima function of the fingers is to prevent conduction in the y-directi while still providing perfect conduction in the z-direction. Co sider a wave whose wavelength, λ_g, in the y-direction is seve times the period of the comb. The structure of the comb averaged out over the wavelength and it behaves like a smoo sheet which conducts perfectly in the z-direction and does n conduct at all in the y-direction. To this degree of approxim tion the comb is uniform in the y-direction and the field bel it depends on y through a factor $\exp(-j2\pi y/\lambda_g)$. Moreov the magnetic field depends on z through the same factor as t current along the fingers, namely $\sin(2\pi z/\lambda)$. Finally, since t field must fall to zero away from the comb, it must depend x through a factor $\exp(+\Gamma x)$, where Γ is the attenuation coe cient in the negative x-direction. Hence, apart from an amp tude factor, we have

$$\left.\begin{aligned} H_x &= \sin(2\pi z/\lambda)\exp[\Gamma x - j2\pi y/\lambda_g] \\ H_y &= B\sin(2\pi z/\lambda)\exp[\Gamma x - j2\pi y/\lambda_g] \end{aligned}\right\} \quad . \quad . \quad (2$$

where B is a constant determined by H_y/H_x.

The attenuation coefficient Γ is determined by the fact th the field components satisfy the wave equation

$$\frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} + \frac{\partial^2 U}{\partial z^2} + \left(\frac{2\pi}{\lambda}\right)^2 U = 0 \quad . \quad . \quad (3$$

By substituting H_x or H_y for U in eqn. (30) we find that

$$\Gamma = 2\pi/\lambda_g \quad . \quad . \quad . \quad . \quad . \quad . \quad (3$$

i.e. the attenuation coefficient away from the guiding stru ture is equal to the phase-change coefficient in the direction propagation.

To evaluate the constant B we invoke Ampère's law: the li integral of the magnetic field around any closed path is equ to $4\pi/c$ times the current threaded through it. The rectangul

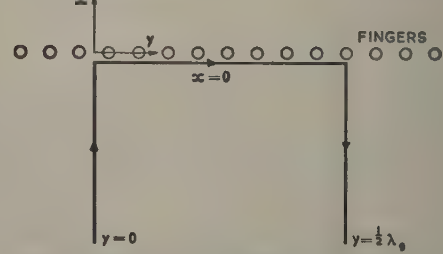


Fig. 9.—Path used in applying Ampère's law.

path in the xy-plane shown in Fig. 9 (which is closed at infini has no current threaded through it, no conduction curre because it is in free space and no displacement current becau $E_z = 0$.

Hence

$$\int_{-\infty}^{0} dx H_x(x, 0, z) + \int_{0}^{\frac{1}{2}\lambda_g} dy H_y(0, y, z) + \int_{0}^{-\infty} dx H_x(x, \tfrac{1}{2}\lambda_g, z) = 0$$

.e. from eqns. (29) and (31),

$$\sin\frac{2\pi z}{\lambda}\left(\frac{\lambda_g}{2\pi} + \frac{B\lambda_g}{j\pi} + \frac{\lambda_g}{2\pi}\right) = 0$$

whence

$$B = -j \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (32)$$

To summarize: the magnetic field distribution below the comb is given by

$$\left.\begin{aligned} H_x &= \sin\frac{2\pi z}{\lambda}\exp\frac{2\pi(x - jy)}{\lambda_g} \\ H_y &= -jH_x \end{aligned}\right\} \quad . \quad . \quad . \quad (33)$$

In our discussion of the interaction of the paramagnetic ions and the signal field we saw that it is convenient to analyse the magnetic field into two contra-rotating circularly polarized components. The component which rotates in the positive sense about $0z$ has magnitude

$$\tfrac{1}{2}|H^+| = \tfrac{1}{2}|H_x + jH_y| = \tfrac{1}{2}|H_x + H_x| = |H_x| \qquad (34)$$

while that which rotates in the negative sense about $0z$ has magnitude

$$\tfrac{1}{2}|H^-| = \tfrac{1}{2}|H_x - jH_y| = \tfrac{1}{2}|H_x - H_x| = 0 \quad . \quad . \quad (35)$$

Thus the magnetic field below the comb has pure circular polarization with a positive sense of rotation about $0z$. On the other hand, the field above the comb ($x > 0$) has pure circular polarization with a negative sense of rotation about $0z$ because, when we pass through the comb, H_y changes sign while H_x remains unaltered.

It was tacitly assumed above that λ_g is positive. By substituting eqn. (33) into eqn. (10) we see that the wave is therefore travelling in the positive y-direction: we call it the forward wave. Consider now a wave travelling in the negative y-direction (the reverse wave). To do this we have merely to change the sign of the coefficient of y in eqn. (29). Following through the above calculation of Γ and B we find that H_x and H_y are both replaced by their complex conjugates. Thus the magnetic field of the reverse wave is the complex conjugate of the magnetic field of the forward wave. For the forward wave we found that $H_y = -jH_x$ below the comb and $H_y = jH_x$ above it. Taking the complex conjugates of these equations we find, for the reverse wave, $H_y = jH_x$ below the comb and $H_y = -jH_x$ above it. Thus the sense of rotation is reversed on both sides of the comb when the direction of propagation is reversed. This is the ideal situation for obtaining non-reciprocal gain.

Suppose we fill the space below the comb with active crystal and subject it to a d.c. magnetic field in the positive z-direction. Then, within the active crystal the magnetic field of the forward wave is rotating in sympathy with the precessional motion, it couples tightly to the paramagnetic ions and the wave is amplified by their stimulated emission. On the other hand, within the active crystal the magnetic field of the reverse wave is rotating in the opposite sense to the precessional motion, it does not couple to the paramagnetic ions and is not amplified.

We can do even better than this by filling the space above the comb with passive crystal (see Fig. 10). The forward wave couples only to the active crystal and is amplified, while the reverse wave couples only to the passive crystal and is attenuated. If the comb is symmetrically loaded with active and passive crystal having the same response to a microwave magnetic field

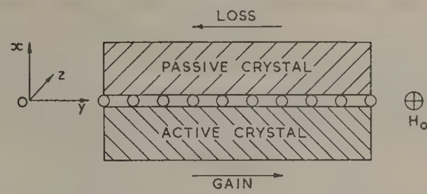


Fig. 10.—The crystal-loaded comb structure.

rotating in sympathy with the precessional motion, the forward gain is equal to the reverse loss. Reflections in the input and output waveguides can never lead to oscillation.

The reader may be in some doubt as to which side of the comb is which. We defined the regions below ($x < 0$) and above ($x > 0$) the comb with reference to the co-ordinate system shown in Fig. 8. By inspecting the Figure and from the above discussion we arrive at the following rule: hold the comb horizontally and look along the fingers in either direction. Then waves travelling to the right rotate in the positive sense (about the direction of viewing) below the comb and rotate in the negative sense above the comb. The directions of rotation are reversed for waves travelling to the left.

(6) ENGINEERING CHARACTERISTICS

The physical principles of the travelling-wave maser have been explained in some detail. In this Section we derive approximate expressions for the important engineering characteristics, i.e. forward gain, reverse loss, bandwidth and noise figure. In so doing we ignore the ohmic loss of the comb structure, because the device will be operating at liquid-helium temperatures where the conductivity of normal metals becomes very high.

First we evaluate the forward gain. A forward wave is amplified because it couples only to the active crystal below the comb. The amplification can be allowed for by multiplying the magnetic field given in eqn. (33) by $\varepsilon^{\alpha y}$ where α is the amplification coefficient in nepers per centimetre. The power flow, P, in the positive y-direction is proportional to the square of the field, i.e. to $\varepsilon^{2\alpha y}$. Hence, in a small length Δy the power flow increased by

$$\Delta P = (dP/dy)\Delta y = (2\alpha P)\Delta y \quad . \quad . \quad . \quad . \quad (36)$$

The increase of P is due to the power emitted by the active ions in the length Δy. Hence

$$\Delta P = P_e\Delta y = \Delta y\iint dxdz(-P_m) \quad . \quad . \quad (37)$$

where P_e is the power emitted per unit length and P_m is the power absorbed per unit volume. The integration is taken over the cross-section of the active crystal. If the active crystal fills the region below the comb in which the field strength is appreciable, the integration limits are $x = -\infty$ to $x = 0$ and $z = 0$ to $z = A$. Equating the two values of ΔP gives

$$\alpha = P_e/2P \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (38)$$

Now the energy stored in the comb travels with the group velocity v_g. Hence

$$P = Wv_g \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (39)$$

where W is the energy stored per unit length.[12] By substituting eqn. (39) into eqn. (38) and using the conventional definition of the Q-factor of the wave, namely

$$1/Q_m = -P_e/\omega W \quad . \quad . \quad . \quad . \quad (40)$$

we obtain

$$\alpha = -\omega/2Q_m v_g \quad . \quad . \quad . \quad . \quad . \quad (41)$$

This is a standard result in the theory of lossy waveguides.[12] In our case, however, the absorbed power is negative, the 'magnetic Q-factor', Q_m, is negative and we have amplification instead of attenuation.

Suppose that the active crystal fills a length l of the comb structure. Then the power flow increases by the factor $\varepsilon^{2\alpha l}$ from one end to the other. Hence the forward gain is

$$G = \varepsilon^{2\alpha l} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (42)$$

and the forward gain in decibels is

$$G' = 10 \log_{10} \varepsilon^{2\alpha l}$$
$$= -27{\cdot}3 \frac{l}{\lambda} \frac{c}{v_g} \frac{1}{Q_m} \quad . \quad . \quad . \quad . \quad . \quad . \quad (43)$$

In the last line we have substituted for α and written $\omega = 2\pi c/\lambda$, where c is the velocity of light and λ is the free-space wavelength at the signal frequency.

To complete the calculation of the forward gain we must evaluate Q_m. The power absorbed per unit volume by the paramagnetic ions is given by eqn. (26); by integrating this equation over the cross-section of the active crystal and assuming that $N(\rho_M - \rho_{M+1})$ and T_2 are constant throughout, we obtain

$$P_e = -\hbar\omega N(\rho_M - \rho_{M+1}) \times \left[(\beta/2\hbar)^2 (S-M)(S+M+1) \frac{2T_2}{1+T_2^2(\omega-\omega_0)^2} \iint dxdz|H^+|^2 \right] \quad (44)$$

The stored energy per unit length is equal to four times the mean magnetic energy stored per unit length below the comb. One factor of 2 comes from the symmetry of the magnetic field and the other appears because the mean stored electric energy per unit length is equal to the mean stored magnetic energy per unit length. [This fact can be verified by calculating, from Maxwell's equations, the electric field going with the magnetic field given in eqn. (33)]. Hence, by making use of eqns. (33) and (34) we obtain

$$W = \frac{4}{8\pi} \mathrm{Av} \left(\iint dxdzH_1^2 \right)$$
$$= \frac{4}{16\pi} \iint dxdz(|H_x|^2 + |H_y|^2)$$
$$= \frac{1}{8\pi} \iint dxdz|H^+|^2 \quad . \quad . \quad . \quad . \quad . \quad . \quad (45)$$

where the limits of integration are the same as before.

When the expressions (44) and (45) are substituted into eqn. (40) the field integrals cancel out, leaving

$$\frac{1}{Q_m} = \frac{2\pi\beta^2}{\hbar} N(\rho_M - \rho_{M+1})(S-M) \times (S+M+1) \frac{2T_2}{1+T_2^2(\omega-\omega_0)^2} \quad (46)$$

Putting in some typical numbers: $S = \frac{3}{2}$, $M = -\frac{1}{2}$, $N = 2 \times 10^{19}$ ions/cm^3, $\rho_{-1/2} - \rho_{+1/2} = -2\%$ and $1/\pi T_2 = 50$ Mc/s gives $Q_m = -95$ when $\omega = \omega_0$. Referring back to eqn. (43) we see that 20–30 dB of forward gain can therefore be obtained from a crystal-loaded comb one wavelength long provided that $c/v_g \simeq 100$. We are helped in this respect by the high refractive index of some paramagnetic crystals. Ruby, for example, has a refractive index of 3·3, so that it automatically reduces the velocity of propagation in the ratio 1 : 3·3. The required extra factor of 1 : 30 is easily obtained with a slightly perturb comb structure. (We neglected the effect of the dielect surroundings when discussing the comb structure in Section The λ which appears there is, in fact, the dielectric waveleng and not the free-space wavelength. The only significant res of this correction, apart from its influence on the group veloci is that the finger length must be one-quarter of the dielect wavelength at the centre of the pass-band).

It is hardly necessary to calculate the reverse loss of travelling-wave maser. A reverse wave is attenuated becau it couples only to the passive crystal above the comb. T magnetic Q-factor of the reverse wave is given by eqn. (46) wh N, T_2 and $\rho_M - \rho_{M+1}$ are given the values appropriate to passive crystal. The density of ions in the passive crystal high enough to ensure that their thermal equilibrium is disturbed by the pump. Now increasing the density of io brings them closer together, they couple to one another mo strongly through their magnetic fields and the paramagne resonance line width increases. Hence, roughly speaking, is proportional to $1/N$ and NT_2 is the same in both active a passive crystals. If we also assume that $\rho_M - \rho_{M+1}$ has same magnitude in both crystals, but opposite sign, then, fr eqn. (46), Q_m has the same magnitude for both forward a reverse waves when $\omega = \omega_0$, but has opposite signs in the t cases. Consequently the reverse loss is equal to the forwa gain. There are, of course, many factors which can alter t conclusion. Equally, there are many ways in which we c adjust the ratio between reverse loss and forward gain.

The amplification bandwidth is easily calculated. It is limi because $-1/Q_m$ falls off as the frequency departs from the cen of the paramagnetic resonance line. The bandwidth betwe the 3 dB points is obtained by doubling the frequency sh necessary to reduce G' by 3 dB from its peak value. The m netic Q-factor can be eliminated from the answer by express it in terms of the peak gain. We find that

$$B = \frac{1}{\pi T_2} \sqrt{\frac{3}{G_p' - 3}} \quad . \quad . \quad . \quad . \quad . \quad . \quad ($$

where G_p' is the peak gain in decibels. Now $1/\pi T_2$ is the l width of the paramagnetic resonance and $\sqrt{[3/(G_p' - 3)]} = 0$ at 20 dB gain. Thus, at 20 dB gain the bandwidth of travelling-wave maser is 42% of the line width. In a cav maser with 20 dB gain the bandwidth cannot exceed 20% the line width and is usually much less.[10]

Finally, there remains the all-important question of no performance. Very little noise arises from the comb structu because it is cooled with liquid helium. This not only gives metal a very low temperature, it also gives it a very low resistiv Nyquist's theorem shows that both these factors reduce the no output. The passive ions also contribute very little to the no because they do not couple to waves travelling towards the o put, and the bulk of the noise output from the cooled parts of maser comes from the active ions.

We shall calculate the noise output in a plausible but rou about way. We neglect noise from the comb structure and passive ions entirely and consider first the case in which 'active' ions are in thermal equilibrium at the standard no temperature, $T_0 = 290°$ K. The 'active' ions are then passi and the forward gain, G, is less than unity. If the impedance the signal generator is also at T_0, the entire device is in therr equilibrium at T_0 and the total noise power output per u bandwidth must be

$$kT_0 = GkT_0 + (1-G)kT_0 \quad . \quad . \quad . \quad ($$

where GkT_0 comes from the source impedance and $(1 - G)kT_0$ from the 'active' ions.

Next we consider the case in which the source impedance is at T_0 while the 'active' ions are in thermal equilibrium at some other temperature T. The temperature of the 'active' ions can be expressed in terms of the fractional population difference across the M to $M + 1$ transition which is excited by the microwave signal. From eqn. (4) we obtain

$$\rho_M^0 - \rho_{M+1}^0 = \frac{E_{M+1} - E_M}{(2S+1)kT}$$

i.e.

$$T = \frac{\hbar\omega}{(2S+1)k(\rho_M^0 - \rho_{M+1}^0)} \quad . \quad . \quad (50)$$

since $E_{M+1} - E_M \simeq \hbar\omega$. Now the noise output from the 'active' ions is proportional to their temperature. Hence the total noise power output per unit bandwidth in this case is

$$GkT_0 + (1 - G)kT \quad . \quad . \quad . \quad . \quad . \quad (51)$$

Finally, we consider the practical case in which the 'active' ions are, in fact, activated by the pump. The populations of levels M and $M + 1$ change to ρ_M and ρ_{M+1}, the 'effective temperature' of the active ions becomes

$$T_m = \frac{\hbar\omega}{(2S+1)k(\rho_M - \rho_{M+1})} \quad . \quad . \quad . \quad (52)$$

which is negative because $\rho_{M+1} > \rho_M$, and the forward gain becomes larger than unity for the same reason. The total noise power output per unit bandwidth in this case is

$$GkT_0 + (1 - G)kT_m \quad . \quad . \quad . \quad . \quad . \quad (53)$$

and we see that the active ions make a positive contribution because $G > 1$ and $T_m < 0$.

The noise figure of the travelling-wave maser is equal to the total noise power output per unit bandwidth divided by the fraction thereof coming from the source impedance, i.e.

$$F = \frac{GkT_0 + (1 - G)kT_m}{GkT_0}$$

$$= 1 + |T_m|/T_0 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (54)$$

since $G \gg 1$ and $T_m < 0$. Substituting some typical values: $\omega = 2\pi \times 10^4$ Mrad/sec, $S = \frac{3}{2}$ and $\rho_M - \rho_{M+1} = -2{\cdot}0\%$ gives $T_m = -6°$K and $F = 1{\cdot}02$, i.e. 0·089 dB. The travelling-wave maser proper has an almost ideal noise performance. The overall noise figure of the amplifier is therefore determined largely by the noise emitted from the uncooled losses in the waveguide input which are amplified by the maser.[13, 14]

(7) CONCLUSION

A complete understanding of masers requires extensive use of quantum mechanics which is not usually familiar to electrical engineers. The semi-classical treatment developed here makes use of only a few quantum-mechanical ideas. The treatment provides a simple introduction to the subject which is essentially correct qualitatively and is not very far from correct quantitatively. It gives an immediate intuitive appreciation of the non-reciprocal behaviour of the paramagnetic ions which is exploited in the travelling-wave maser.

The application of the semi-classical theory has been confined to the travelling-wave maser because the cavity maser has been discussed extensively in the references quoted. The travelling-wave maser is superior to the cavity maser in many respects. Wide bandwidths are possible because of the non-regenerative nature of the amplification. Loss in the reverse direction can be built into the device by properly locating passive maser crystals or passive ferrimagnetic crystals.[4] The reverse loss ensures that the gain is relatively insensitive to changes in the load. The travelling-wave maser is easy to tune and saturates at a higher power level than the cavity maser.[10] Both the cavity maser and the travelling-wave maser have noise figures of the order of 0·1 dB. An intensive development of the travelling-wave maser is to be expected in the next few years.

(8) REFERENCES

(1) Wittke, J. P.: 'Molecular Amplification and Generation of Microwaves', *Proceedings of the Institute of Radio Engineers*, 1957, **45**, p. 291.

(2) Bloembergen, N.: 'Proposal for a New Type Solid-State Maser', *Physical Review*, 1956, **104**, p. 324.

(3) Weber, J.: 'Masers', *Reviews of Modern Physics*, 1959, **31**, p. 681.

(4) DeGrasse, R. W., Shulz-Dubois, E. O., and Scovil, H. E. D.: 'The Three-Level Solid State Travelling-Wave Maser', *Bell System Technical Journal*, 1959, **38**, p. 305.

(5) Richtmeyer, F. K., Kennard, E. H., and Lauritsen, T.: 'Introduction to Modern Physics' (McGraw-Hill, 1955), Chapter 7.

(6) Bleaney, B., and Stevens, K. W. H.: 'Paramagnetic Resonance', *Reports on Progress in Physics*, 1953, **16**, p. 107.

(7) Ditchfield, C. R., and Forrester, P. A.: 'Maser Action in the Region of 60° K', *Physical Review Letters*, 1958, **1**, p. 448.

(8) Butcher, P. N.: 'Theory of Three-Level Paramagnetic Masers: Part 1—Quantum Theory', *Proceedings I.E.E.*, Paper No. 2641 R, May, 1958 (**105** B, Supplement No. 11, p. 684).

(9) Pake, G. E.: 'Nuclear Magnetic Resonance', from Vol. 2 of 'Solid State Physics' (Academic Press, 1956).

(10) Butcher, P. N.: 'Theory of Three-Level Paramagnetic Masers: Part 2—Amplification and Oscillation', *Proceedings I.E.E.*, Paper No. 2642 R, May, 1958 (**105** B, Supplement No. 11, p. 691).

(11) Butcher, P. N.: 'The Coupling Impedance of Tape Structures', *ibid.*, Paper No. 2295 R, March, 1957 (**104** B, p. 177).

(12) Slater, J. C.: 'Microwave Electronics' (Van Nostrand, 1950).

(13) Butcher, P. N.: 'Theory of Three-Level Paramagnetic Masers: Part 3—Output Noise-Power Spectrum', *Proceedings I.E.E.*, Paper No. 2643 R, May, 1958 (**105** B, Supplement No. 11, p. 699).

(14) Butcher, P. N.: 'Theory of Three-Level Paramagnetic Masers: Part 4—Noise Figure', *ibid.*, Paper No. 2644 R, May, 1958 (**105** B, Supplement No. 11, p. 705).

[The discussion on the above paper will be found overleaf.]

DISCUSSION BEFORE THE ELECTRONICS AND COMMUNICATIONS SECTION, 7TH MARCH, 1960

In opening the discussion **Dr. J. C. Walling** drew attention to the fact that a comb-type circuit of the type used in the travelling-wave maser produced elliptically polarized waves in the region where the active material was located, and that, in addition, the side walls which gave the desired dispersion characteristic resulted in a linearly polarized component. The author agreed with this and commented that it was necessary to avoid interference with correct operation by designing the structure so that the amplitude of the unwanted components was small relative to the wanted circular one.

Gain Calculation.—A matter which led to some detailed discussion between the author and **Dr. Walling, Mr. L. Lewin** and **Dr. W. E. Waters** was that of the correct expression for calculation of the gain. The author pointed out that his results were derived directly from the resistive-loaded-waveguide case by replacing positive by negative resistance. **Mr. Lewin** supported the author's conclusion by referring to the effect of the phase of excitation of a distributed negative resistance on power in forward and backward directions, and concluding with the simple example of two sources placed so that their equal electric fields added in the forward direction and subtracted in the backward. This point of view was supported by **Mr. Waters.**

Bandwidth.—Questions relating to bandwidth were raised by **Dr. Waters** and **Mr. W. E. Willshaw.** The reasons for the need for a large circuit bandwidth relative to the much smaller electronic bandwidth were explained by the author, who pointed out that this was to accommodate the wide tuning range available from magnetic-field variation. On the question of possibilities for significant increase of amplifier electronic bandwidth the author suggested that an increase of a few times, say to 100 Mc/s, was possible by grading the magnetic field along the length of the amplifier, so that the frequency for maximum gain varied and the bandwidth was thus increased. A useful increase of this kind was, of course, dependent on use of minimum group velocity for the circuit, so that maximum gain per unit length of amplifier was obtained.

In answer to the question why the electronic bandwidth was limited to 42% of the line bandwidth of the material, the author pointed out that this was the maximum value which could be used without the overall gain falling by more than 3 dB at the band edges. This was, of course, a smaller figure than that for the material bandwidth, for which the reciprocal Q-factor of the material falls by 3 dB at the band edges.

Finally, in answer to a question by **Mr. Willshaw** on the possibility of increasing the inherent electronic bandwidth of the material, for example by increase of coupling between ions so that each main energy level made use of has a greatly increased bandwidth, the author thought that this was not very likely, and he did not know of any research in progress on this subject. The difficulty was to get sufficient gain even at present bandwidths of materials, and the first way to increase overall bandwidth was probably to use a spatially varying magnetic field, as already mentioned.

Pump Power.—**Mr. Lewin** raised a number of questions about the attenuation of the comb structure to the pump power and the pump power required. The author said that the structure attenuation was about 3 dB and that he thought the power absorbed in the maser material was about 10 mW, although only about 1–2 mW might be used in the amplifying material, the remainder being used in the heavily doped material used for the non-reciprocal attenuator. In answer to a question as to whether resonating of the circuit at the pump frequency would result in inactivity of parts of the crystal at the nodes of the magnetic field, the author suggested that this might be avoided by arranging for the material to be oriented so that, even though it might be at a null of the transverse field, it coupled to the longitudinal component of the field.

Crystal Orientation.—The subject of crystal orientation was raised by **Mr. L. Birch,** who queried whether the crystal axis needed to be critically set to the magnetic field to achieve the required energy differences. The author commented that it was necessary to cut and orient the crystal carefully, and in particular make sure that the chromium ions looked like dipoles so far as the signal was concerned: they experienced strong perturbations, owing to the electric fields of the surrounding crystal.

Mr. Birch asked whether the energy levels in chromium, which has a complicated internal structure, were calculable, and it was stated that it was possible to predict to within 1% the energy levels for a particular direction of magnetic field and to cut the crystal accordingly.

Noise.—The aspect of noise performance led to a number of queries and comments. In answer to a question from **Mr. J. L. Eaton** the author outlined some of the factors involved. With an amplifier of effective temperature lying in the liquid-helium range a noise figure in the range 0·1 dB was to be expected and a fundamental factor affecting this would be the temperature of ions excited by the pump as affected by the disturbance of equilibrium.

Dr. Walling commented that noise introduced by the necessary connection of finite loss between a room-temperature source and the input to the amplifier could have a major effect on overall amplifier noise, being 2–3 times bigger than that due to the maser crystal itself.

The question of how low the operating temperature needed to be was raised by **Dr. Waters,** having in mind the use of liquid hydrogen rather than liquid helium. In reply the author pointed out that the need for the low temperature arose from the necessity to reduce the thermal vibrations of the crystal with which the pump power had to compete. Another trouble with the higher temperatures was that the difference between populations at the various levels was reduced so that, with too high a temperature, the amplification process was ineffective. However, cavity-type (narrow-band) masers had been operated up to dry-ice temperatures. In travelling-wave masers it was possible to use liquid hydrogen, but although liquid nitrogen could be considered, the gain would be low, particularly since the copper losses of the circuit would be increased at the higher temperature.

Dr. P. Vigoureux asked whether, since the essence of low noise performance in the maser was low temperature, the same technique could not be applied to transistor and other amplifiers. The author pointed out that the effect of cooling in the transistor would be deleterious, owing to freezing out of the carriers. However, he saw no reason why parametric amplifiers should not be cooled down, certainly to the liquid-nitrogen range, with performance rivalling the maser.

Applications.—Turning to the use of masers, **Mr. G. Millington** asked for information about the field of applications and possible trends of performance over the next few years. The author emphasized that the maser was of interest only in fields where a noise factor of 0·1 dB was important. Applications of the cavity maser had so far been in the field of radio-astronomy, in seeking microwave emission from Venus and radar echoes from the Moon—in the latter case with a 10:1 improvement of sensitivity. Its use in a satellite was hard to foresee; closed-circuit refrigeration systems would be essential. The travelling-wave maser had been used to measure sky temperatures. In answer to questions raised by **Mr. R. J. Halsey** in connection with its use in systems involving reflection from satellites, the author com-

nented that it might be worth using a maser if the receiver aerial vas focused above the horizontal and was properly designed. The usual type of aerial used to-day would not suffice, owing to he collection of a large amount of noise power from the ground. The point was that the maser is so good that all the other components need to be redesigned around it.

The question of cost of maser material was also raised by Mr. Halsey, but it was pointed out that the material used is ommercial ruby and is not expensive.

Figure of Merit.—Finally, **Mr. D. W. L. Dickinson** raised the question whether it was possible to apply some figure of merit to maser performance, so that the optimum compromise between the limits involved in the different variables could be achieved. The author felt that at the present state of knowledge there was not enough information to do this, and a number of physical parameters could not yet be determined. One thing which was known was that it is of value to increase the pump frequency, since the maximum change of ion population can thereby be achieved. However, in spite of all the gaps in our knowledge at the present time, the travelling-wave maser was 'quite a device'.

621.317.335.3: 621.315.612

The Institution of Electrical Engineers
Paper No. 3223 M
July 1960

DETERMINATION OF THE DIELECTRIC PROPERTIES OF LOW-LOSS CERAMICS AT Q-BAND FREQUENCIES

By J. M. FREE, B.Sc., Student, and G. B. WALKER, Ph.D., M.A., Associate Member.

(*The paper was first received* 21*st July*, 1959, *and in revised form* 6*th January*, 1960.)

SUMMARY

Although the difficulties involved in the measurement of dielectric properties increase with frequency, a conventional cavity method with minor modifications has proved satisfactory at millimetre wavelengths.

A cavity, ½ in in diameter, was excited in an H_{01} mode by means of coupling holes. It was found to have sufficiently low wall loss to enable the loss tangents of disc specimens of magnesium-titanate and titanium-dioxide ceramics to be measured. Both loss tangents were of the order of 0·0003, and the relative permittivities were 14 and 80, respectively.

Although a high degree of accuracy is not claimed, useful results can be obtained with relatively simple equipment. A convenient method of marking small frequency intervals is described.

(1) INTRODUCTION

For many applications of millimetre waves the dielectric properties of low-loss ceramic specimens is of the utmost interest. A relatively simple method of determining these properties was desired and it was decided to investigate whether conventional cavity methods could provide worth-while results at frequencies as high as 35 Gc/s.

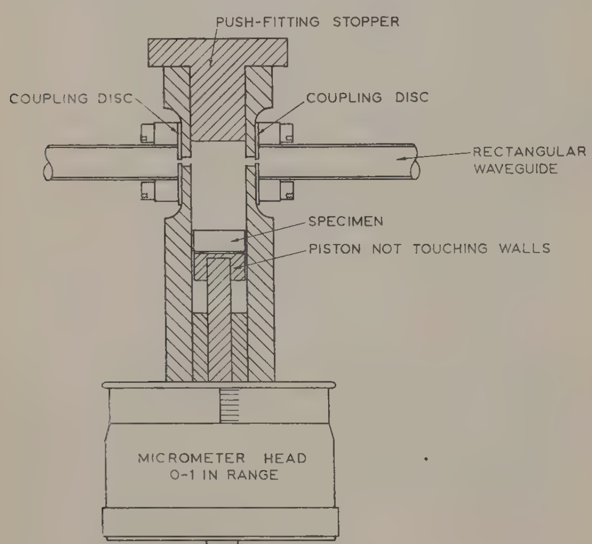


Fig. 1.—Resonant cavity.

The principal difficulty with most methods is the increase of wall loss with frequency. This can be largely avoided by using the H_{01} mode in a cylindrical cavity, and it was found that very sharp resonances ($Q \sim 15\,000$) could be obtained. Other modes were only feebly excited. Mechanical tolerances and the availability of power caused much less difficulty than had been expected. As the H_{01} mode pattern is circularly symmetric, the troublesome effects due to double degeneracy often found with other modes are avoided.

The cavity used is shown in Fig. 1. The specimens take the form of discs several wavelengths thick, located at one end of the cavity.

(2) OUTLINE OF METHOD

The relative permittivity is calculated from the change, Δ, in cavity length needed to restore resonance on inserting the specimen.* This is found for all observed modes (H_{01}, H_{11}, H_{21}, H_{31}). One of the difficulties in the experimental work is distinguishing between the resonances of different modes. All the resonances are known to have been identified correctly, for the value of Δ for each indicates the same relative permittivity for the specimen. Fig. 2 demonstrates that the resonances have been identified correctly, for, wherever a resonance of a particular mode is observed, its presence can be predicted from theory once the relative permittivity is known.

To determine the relative permittivity of the specimen, one must solve the equation:

$$\frac{\tan \beta_s d}{\beta_s d} = \frac{\tan \beta_a(\Delta + d)}{\beta_a d}$$

where β_s and β_a are the propagation coefficients in specimen and air and d is the specimen thickness. As the specimen is several wavelengths thick, there are several values of β_s, and hence of relative permittivity, which could satisfy this equation. By repeating the experiment with a different-size specimen (it was found sufficient to grind off 10 mils) the unique solution can be obtained.

The loss tangents, defined as $\sigma/\omega\epsilon$, where σ is effective conductivity, ω angular frequency and ϵ permittivity, is calculated from the observed Q-factors of the H_{01} mode resonances. Following the method given by Barlow and Cullen,* allowance is made for the losses in the walls of the cavity (which were about half the losses in the specimen) by measuring the Q-factors of the air-filled cavity and hence calculating what the Q-factor will be if there is no loss in the specimen. Errors will arise in this calculation if the wall loss is concentrated at some imperfection or in some piece of foreign matter which may have entered the cavity. The latter is unlikely since the measured Q-factor of the air-filled cavity did not change greatly (7%) over three months, during which time it was dismantled and cleaned several times. The former would cause the loss tangents, as calculated from experimental measurements taken at successive resonances of the H_{01} mode, to be inconsistent. The measured Q-factors of the air-filled cavity at successive resonances would not increase in the ratios which could be predicted from theory by comparing

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.

Mr. Free and Dr. Walker were formerly in the Department of Electrical Engineering, Queen Mary College, University of London. Mr. Free is now in the Engineering Laboratory, University of Oxford, and Dr. Walker is now Research Professor at the University of British Columbia, Vancouver.

* BARLOW, H. M., and CULLEN, A. L.: 'Microwave Measurements' (Constable, 1950), p. 283.

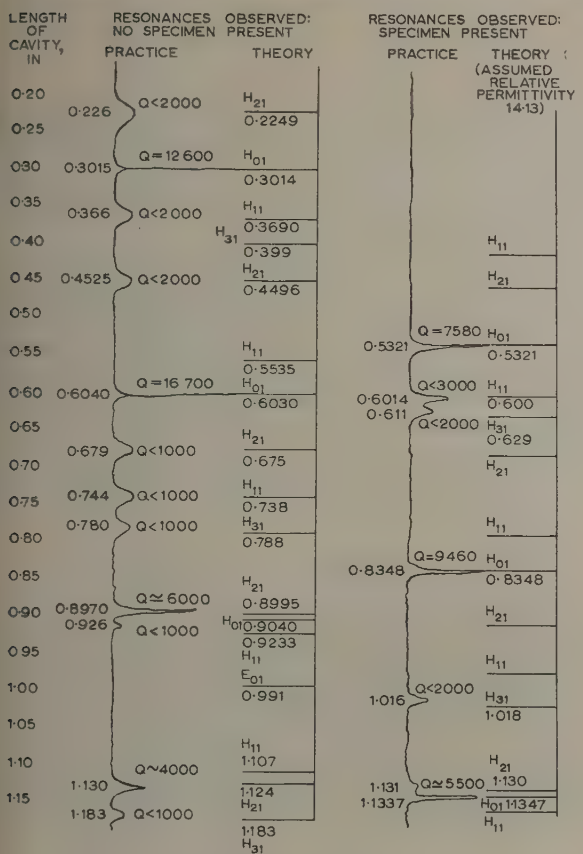


Fig. 2.—Resonances in the cavity with and without a magnesium titanate disc.

A typical diagram showing how the various modes can be readily identified by comparing the observed resonances with theory. The Q-factors and resonant lengths are marked.

the loss on the curved walls with that on the end walls. In fact, the loss tangents indicated by three successive resonances were mutually compatible. The ratio of the Q-factor of the first resonance to that of the second was in agreement with theory within 3%. That of the third resonance could not be measured accurately owing to the coincidence of the H_{01} and the H_{21} modes at almost the same cavity length (see Fig. 2).

(3) CAVITY RESONANCE TECHNIQUE

(3.1) Resonant Cavity

The resonant cavity is a brass tube of $\frac{1}{2}$ in bore, closed at one end by a push-fit stopper and at the other by a piston attached to the shaft of a sensitive micrometer (see Fig. 1). The piston is arranged not to touch the walls of the tube. This has negligible effect on the H_{01} mode so long as the clearance is small, but causes a large damping of E modes. The piston travel of 1 in is sufficient to obtain three H_{01} mode resonances to provide cross checks on results. Power is coupled into and out of the cavity by small holes one quarter wavelength from the push-fit stopper. There is further control of the power by thin discs with smaller central holes which are clamped under the flanges where the waveguide is bolted to the cavity. The broad face of the waveguide lies parallel to the axis of the cavity. This, together with the fact that the coupling holes are one quarter the H_{01} mode wavelength from one end of the cavity, ensures that the H_{01} mode is launched in preference to other modes. No E mode resonance could be observed: the H_{11}, H_{21} and H_{31} resonances were observed only with extreme care.

The optimum size of the coupling holes was determined empirically. If too large, it was possible to change the measured Q-factor of the cavity by external adjustments (e.g. to the attenuators or matching stubs). If too small, the response of the cavity was obscured by noise. It was found that as the input-power hole was decreased in size, the measured Q-factor increased less and less rapidly and beyond a certain point no significant increase occurred: similarly, with the output hole. A hole size of about $\frac{1}{16}$ in was found to give a reasonable output level with no perceptible (3%) change in the Q-factor by any adjustment external to the cavity.

All internal surfaces of the cavity were silver plated after machining. The surface finish must be of high quality because, unless the losses can be kept low, the increase of loss on introducing the specimen is too small for the results to be accurate.

The cavity when air filled gave high Q-factors in the range 10 000–17 000 for the H_{01} mode resonances. The measured Q-factor varied less than 3% when the attenuation between the klystron and the cavity was varied over a range of 18 dB and during two hours of test. In the three months prior to the testing of the specimens the Q-factor remained substantially constant, although the cavity was dismantled and reassembled several times.

(3.2) Measurement of Q-factor

Resonance curves were displayed on one trace of a double-beam oscillograph by frequency modulating the klystron from

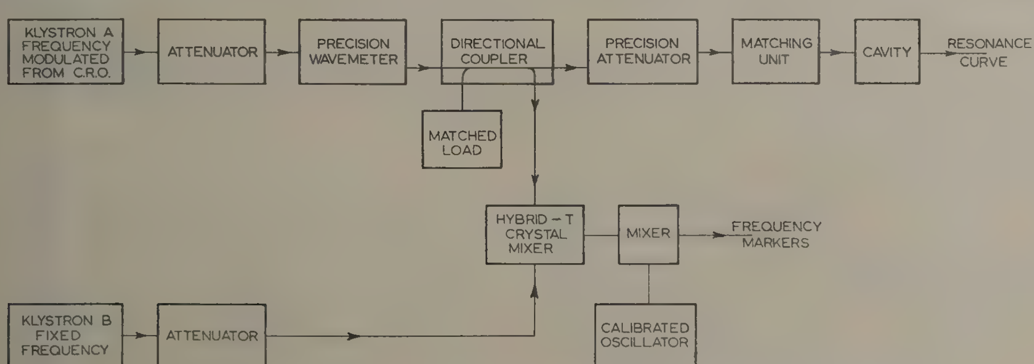


Fig. 3.—Block schematic of apparatus for measuring Q-factors.

the oscillograph time-base. The Q-factor of each resonance was measured from the frequency difference of half-power points in the conventional way.

The second trace was used to display marker pips at known frequency intervals. This was done using a rather simple circuit, yet one which can produce marker pips at any desired frequency intervals up to hundreds of megacycles per second. The sharpness and size of the pips can be varied as required, so that pips of a few kilocycles per second apart can be readily distinguished.

In the circuit of Fig. 3, the klystron output is first mixed with that of a second klystron, which is fixed in frequency, by means of a hybrid-T junction and crystal. As super-high-frequency components are lost, a frequency-modulated signal, swept either side of zero frequency, appears in the crystal output. When this is mixed with the output of a calibrated oscillator and filtered, three marker pips appear. The centre pip is produced when the two klystrons are in tune and the outer two when they differ by the oscillator frequency. The mixer output is then amplified by a transistor audio amplifier. Marker pips of any desired size and sharpness can be produced by variation of the power input from the second klystron, the power from the calibrated oscillator and the gain of the amplifier. (It was not found necessary to alter the amplifier frequency response.)

The effect of harmonics of the oscillator and klystron frequencies can be detected by the appearance of other marker pips when the klystron frequencies differ by half, twice, thrice, etc., the oscillator frequency. These cause no confusion as they are so much smaller than the others. That the largest pips appear when the klystron frequencies differ by the oscillator frequency has been demonstrated by a precision wavemeter (e.g. when the oscillator was set at 50 Mc/s the wavemeter indicated that the large outer marker pips were separated by 100 Mc/s).

The wavemeter pip can be made to appear on the same trace as the frequency markers. Thus both changes in frequency and absolute frequency are known.

(4) DIELECTRIC SPECIMENS

The specimens rested on the micrometer piston and were sliding fits in the cavity bore. The dielectric losses of these specimens were small; thus specimens several wavelengths thick had to be used to make the specimen losses greater than the wall losses. All specimens were originally ground to $\frac{1}{2}$ in in diameter and $\frac{1}{4}$ in thick using a diamond wheel. These were tested and then ground down until the dielectric losses were reasonable. The thickness is critical, particularly if near an integral number of half wavelengths. Great care was taken during the final grinding, which was performed by hand using fine carborundum powder to ensure uniformity of thickness.

It was found, as suspected, that specimens with perceptible degrees of variation of thickness gave completely misleading Q-factors. For example, one with a thickness varying from 0·250 to 0·252 in across its diameter gave Q-factors of less than 1 000 at each of the three H_{01} mode resonances. The same specimen ground down to 0·238 8 in with no perceptible thickness variation gave Q-factors of 7 500, 9 000 and 5 500 at the three resonances. The wavelength in this specimen was 0·09 in. The two specimen sizes, however, gave values for relative permittivity agreeing within one part in 400.

(5) RESULTS

The frequency was 34·8 Gc/s in all measurements.

(5.1) Magnesium-Titanate Specimen

The relative permittivity of the magnesium-titanate specime was found to be 14·13. The maximum error likely in th determination is $\frac{1}{2}$%. When the specimen was ground dow from 0·251 to 0·238 8 in in thickness the relative permittivi agreed within one part in 400, although the cavity had mea while been dismantled and reassembled.

The loss tangent was estimated to be $3 \cdot 1 \times 10^{-4} \pm 30$% The errors are due to the uncertainties in specimen thicknes position in the cavity, etc., and not to the Q-factor measurement The loss in the specimen was slightly greater than that in th walls of the cavity. The reduction of Q-factor at the fir resonance indicated a loss tangent of $2 \cdot 9 \times 10^{-4}$, the secon one of $3 \cdot 9 \times 10^{-4}$ and the third, one of $2 \cdot 9 \times 10^{-4}$. The la result is unreliable as previously discussed, owing to the coinc dence of the modes.

The same specimen, after drying out in an oven at 200° for four hours, showed no significant change in either relati permittivity or loss.

(5.2) Titanium-Dioxide Specimens

The relative permittivity of one specimen of titanium dioxi (0·111 5 in thick) was found to be 80. The maximum like error is 2%. A thicker specimen (0·252 3 in) had a relati permittivity of 76·5 ± 2%.

The loss tangent of the thinner specimen was estimated to $3 \cdot 3 \times 10^{-4}$ and definitely neither more than 7×10^{-4} nor le than $2 \cdot 4 \times 10^{-4}$. The effect of any surface irregularity in th case is the cause of the uncertainty in the estimation of the los Defects are very serious, for the wavelength in the specimen only about one millimetre.

(6) CONCLUSIONS

The work has shown that, owing to the unique properties the H_{01} mode, a cylindrical cavity with low loss can be realiz at Q-band frequencies. With this the relative permittivity (if le than 100) and loss of any material can be determined. F high-loss materials, specimens a fraction of a wavelength thi must be used: for low loss, larger specimens must be used.

The double heterodyne method of measuring frequency inte vals has been shown to be simple to set up and flexible operation.

A specimen not precisely uniform in thickness has been show to give a relative permittivity closely in agreement with that of more accurately ground specimen, but a loss tangent whic would be most misleading. Similarly, it is suspected that th surface condition of a specimen may upset the measurement the loss tangent by exerting some control on the energy stored the specimen. This would not noticeably affect the relativ permittivity measurement, unless the wavelength in the specime were minute.

(7) ACKNOWLEDGMENTS

This work, carried out in the Department of Electric Engineering, Queen Mary College, University of Londo formed part of a project supported by the Admiralty and th D.S.I.R. The ceramic specimens were specially prepared b Dr. E. Shaefer.

621.319.4

The Institution of Electrical Engineers
Paper No. 3261 M
July 1960

SOME MECHANISMS OF FAILURE OF CAPACITORS WITH MICA DIELECTRICS

By A. A. NEW, M.Sc., F.R.I.C., A.Inst.P.

(*The paper was first received* 22*nd September, and in revised form* 3*rd November*, 1959.)

SUMMARY

During investigations of mica capacitor failures of various equipments during the last ten years and studies to improve their reliability, many mechanisms of failure have been examined. The principal features of the mechanisms are described and illustrated, and are summarized in tabular form for quick reference. Some methods of examination and dissection of these capacitors with the minimum loss of evidence are given in detail.

This work does not imply that the proportion of mica capacitors which fail in service is excessive. Some of the causes of failure would occur in other types of capacitor, perhaps to a similar extent.

(1) INTRODUCTION

Mica capacitors made from the best quality mica under suitable well-controlled conditions are generally reliable and stable, but under other conditions failures are liable to occur.

During a number of investigations of causes of failure of certain of these capacitors during the last ten years, particularly in line telecommunication equipment, and in amplifiers intended for submerged repeaters over the years 1954–58, it was necessary to augment the scattered published information with experimental work on the exact appearance and characteristics of faults arising in mica capacitors.

This summary of the work, which, with extracts from the literature, forms Sections 2, 3 and 4, is intended to simplify any similar investigations in future and to assist in improving the reliability and stability of mica and similar types of capacitor.

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.
Mr. New is at the Post Office Research Station.

(2) CAUSES OF SHORT- AND OPEN-CIRCUITS

(2.1) Puncture of Dielectric due to Excessive Voltage

The intrinsic electric strength of Muscovite mica is usually stated to be about 10×10^6 volts/cm for thicknesses in the range 0·02–0·07 mm, corresponding to 25 000 volts/mil. Nevertheless, when laminations of good-quality mica are subjected to an increasing potential, it is generally found that they break down at voltages of the order of one-fifth of this. Manufacturers generally allow a safety factor of 10–20 on the latter figure in assigning a continuous working voltage for capacitors made from the mica, to cover random variations in strength.

A number of capacitors that had either broken down under known conditions of service or been deliberately broken down under controlled conditions were dissected as described in Section 8 and examined. The results and other relevant data are given in Table 1, and photographs of the seat of breakdown in two of them are shown in Figs. 1 and 2. From consideration of the data and photographs it was apparent that the following features are frequently associated with breakdown due to excess voltage:

(*a*) A small puncture in one or more plates.
(*b*) About five to eight cracks radiating from the main puncture.
(*c*) Location of the puncture at a region of high electrical stress, particularly at the edge or corner of an electrode.
(*d*) In a stack, adjacent plates have indentations (and perhaps cracks) in exact register with the puncture, which extend through the stack with decreasing intensity to an extent which is related roughly to a fractional power of the energy dissipated at breakdown and to the rigidity of the capacitor. Occasionally the marks may terminate abruptly at a very thick plate (e.g. where a cover plate three to ten times the thickness of the laminations has been inserted).

Table 1

Some Breakdown Characteristics

Protection and impregnation	Capacitance (numbers in brackets are externally added capacitances)	Mica plates		Breakdown voltage	Charge	Broken down mica, sheet number	Point of breakdown	Number of marked micas each side of breakdown	Insulation resistance after breakdown
		Number	Thickness						
	pF		in	kV	µC				kΩ
Wax-coated and wax-impregnated	3 300	3	0·001 8	5·4	18	2	C	*1, 1*	$>10^7$
	3 300(+3 300)	3	0·001 8	4·0	26	3	B	*2, 0*	$>10^7$
	3 300(+6 600)	3	0·001 8	3·0	30	3	C	1, 0*	$>10^7$
	30 000	34	0·001 3	about 0·1	3	15	B	0, 0	60→12
Encapsulated in epoxy resin and oil-impregnated	4 600	23	0·001 6	3·5	16	21	E	6, 4*	140
	5 700	26	0·001 4	5·5	31	7	EC	7, 3*	500
	10 500	49	0·001 5	4·5	47	43	EC	9, 6*	400
	15 000(+ 6 000)	71	0·001 5	4·2	63(88)	33	EC	8, 5*	
	37 000(+68 000)	116	0·001 7	0·75	28(78)	8 and 9	B	0, 0	12→50 (9 volts)
	33 895(+2 000)	95	0·001 5	0·38 a.c. (twice)		21	E	12, 12	100→140→5·2

C = Corner of plate.
E = Edge of plate.
B = Body of plate.
* The next plate in the stack was an extra thick one not carrying electrodes.

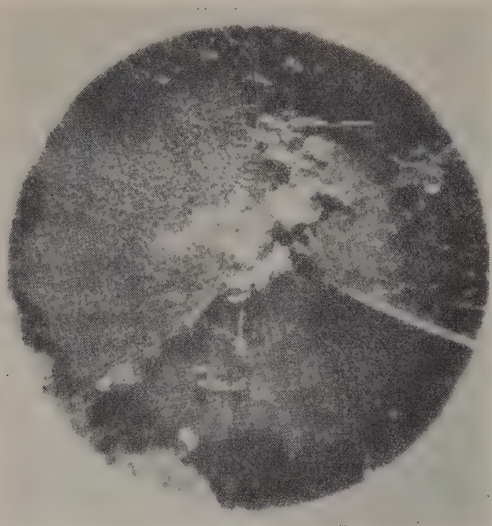


Fig. 1.—Enlarged view of point of breakdown due to excessive voltage.
Magnification, ×33.

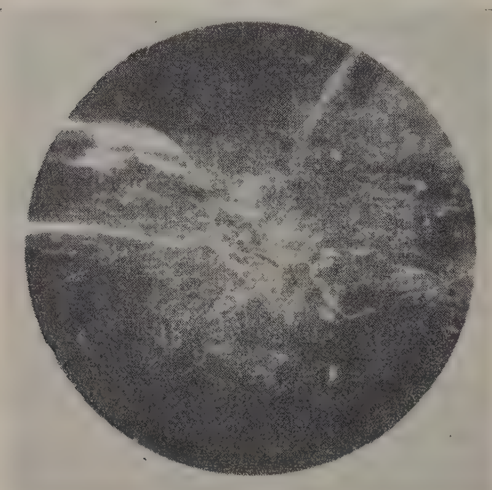


Fig. 2.—Enlarged view of point of breakdown at a few hundred volts, due to a fault in the mica.
Magnification, ×33.

(2.2) Surface Flashover or Breakdown due to Excessive Voltage

Under normal conditions when the surface of the mica is clean, initial flashover is almost entirely a gaseous phenomenon.[3] To estimate the clearances necessary to avoid flashover a knowledge of the breakdown potential is required. The breakdown voltage between surfaces of the order of 1 cm radius separated by 2 mm is about 8 kV in air under ordinary conditions; with a radius of 0·1 mm the value[1, 2] is about 3 kV, and extrapolating to a radius of 0·01 mm (half the thickness of the electrode of a silvered mica capacitor), about 2 kV. Since mica capacitors are usually given a short proof test at three to five times their working voltage, the spacing between the edges of electrodes is usually made not less than 2 mm (0·08 in).

In some cases the tendency to flashover may be increased by the presence of very thin and irregular deposits of silver in the nominally clear space outside the electrodes, caused by faulty printing or spraying of the silver. If oil, wax or other organic matter is present, a discharge may decompose some of it to carbon. This may lead to branched markings of carbonized material known as tracking[3] or 'treeing'.

The commonly-occurring symptoms of flashover, apart from the spark or discharge pulse, are the absence of any puncture in the dielectric and, if enough charge has been dissipated, a series of characteristic markings (see Fig. 3) across the surface

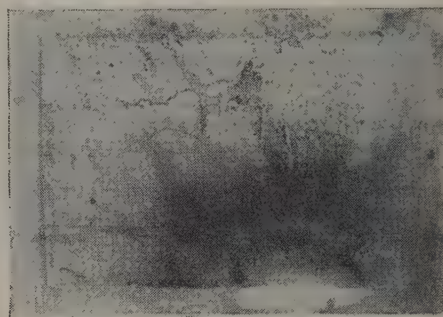


Fig. 3.—Silvered mica on which flashover has taken place.
Magnification, ×2.

of the dielectric joining the points where flashover took place. These may be of sputtered electrode metal, or may be black and perhaps of tree-like form if any organic matter is present to lead to charring or tracking. It follows from the first two paragraphs that the location of the fault will be at the point of lowest electric strength through air, which will depend partly on nearness of approach of parts at different potentials and the effect of their shape on the field strength. Experimental data on some capacitors that flashed over are given in Table 2.

(2.3) Puncture or Damage due to Enclosed Particles

Adventitious particles may cause damage if they are harder than the mica and of dimensions greater than about a quarter the thickness of the mica plates. The likeliest origins of such particles are grit from industrial atmospheres, fragments of wire, filings, solder, etc., or sand from the weathering of mortar. Failure may occur by the particle making a partial or complete puncture of a plate when a stack of plates is pressed together, or, in the case of metal particles, by forming a short-circuit.

In some cases the appearance after dissection resembles that of a capacitor that has failed from excessive potential, particularly if, as is often the case, the particle itself has been destroyed.

The majority of such cases are eliminated during the proof testing, but partly-punctured plates may withstand the proof test and yet fail later under working voltage. Under some circumstances, where direct evidence of the conditions at the time of failure is not available or is in doubt, it may be important to determine by subsequent examination whether excessive voltage or an enclosed particle was the cause of the fault.

Three lines of experimental work were followed: detection of particles by microscopic examination, detection of particles by radiography, and microscopic examination of the mechanical damage.

(2.3.1) Detection of Particles by Microscopic Examination.

There is a wide range of types of construction of mica capacitors. It is possible with care to dissect the simpler ones with sufficiently little disturbance to find any particles which may be

Table 2

SOME FLASHOVER CHARACTERISTICS

Protection and impregnation	Capacitance	Mica plates		Flashover voltage	Charge	Flashover distance	Insulation resistance after flashover
		Number	Thickness				
	pF		in	kV	μC	in	MΩ × 10³
Wax, wax-impregnated	1 000	1	0·001 7	8·1	8·1	0·12	5·0
	1 000	1	0·001 7	6	6·0	0·09	60
	3 300	3	0·001 8	4	13·2	0·06	44
	5 600	5	0·001 8	1·9	10·6	0·10	220
	5 600	5	0·001 8	2·2	12·3	0·08	250
	5 600	5	0·001 8	2·2	12·3	0·10	230
	220	1	0·001 8	2·9	0·6	0·08	1 300
	220	1	0·001 8	1·4	0·3	0·05	1 500
	220	1	0·001 8	2·1	0·4	0·06	1 000
	100(+200)	1	0·001 8	7·5	2·3	0·05	60
	100(+100)	1	0·001 8	7·5	1·5	0·06	60
	100	1	0·001 8	9	0·9	0·06	60
Phenolic resin, wax-impregnated	3 300(+6 600)	10	0·001 7	6	59·0	0·08	450
	6 200	18	0·001 8	3·3	20·5	0·05	0·37
	6 200	18	0·001 8	2·5	15·5	0·07	140
	6 200	18	0·001 8	5·0	31·0	0·08	100
	220	4	0·001 8	2·9	0·6	0·04	2 000

present, but the difficulty increases with the number of plates in the stack. Methods for removing various types of casing or encapsulation are given in Sections 8.1 to 8.3. In processes for making silvered-mica capacitors where the plates are given a second firing in stack form they often become firmly bonded together and it is impossible to separate them satisfactorily as silvered-mica plates. By an alkaline treatment they can sometimes be separated fairly well into micas and detached silver electrodes, and by an acid treatment the micas alone can be separated excellently. However, the first process would eliminate all particles soluble in alkali, and the second, all of those soluble in acid. The processes are described in Sections 8.4 and 8.5.

(2.3.2) Detection of Particles by Radiography.

Since the degree of stopping of X-rays by a capacitor is a function of the thicknesses and densities of its constituents, there must be a range of conditions within which embedded particles can be detected. To determine whether this range includes conditions likely to be met in practice, some 10-plate silvered-mica capacitors were prized open and small fragments of fine wires of known diameters were inserted, after which the capacitors were closed and X-rayed under various conditions.

In one of these capacitors (which have thick mica cover plates at each end of the stack) there were inserted ten short pieces of copper wire of diameters 0·006, 0·004, 0·002 5, 0·001, 0·000 5, 0·000 5, 0·001, 0·002 5, 0·004 and 0·006 in, in that order. Each piece of wire was tightly knotted with a single knot in the middle, representing a particle of about twice the diameter of the wire. In the original negative (sideways view) even the 0·000 5 in diameter wire can be clearly distinguished in the optimum exposure, but the knots in the 0·000 5 in wire are only just detectable, suggesting that particles of metal as dense as copper (density 8·9 g/cm^3) would be detectable down to about 0·001 in under these conditions, but that thinner sections than this would be visible only if the particle extended several thousandths of an inch in at least one direction. This only applies when using the optimum back illumination and a magnification of about ×10. With an ordinary naked-eye examination, the limit is probably nearer 0·003 in. Some of the larger wires could be seen in two of the radiographs taken edgewise, but this was largely a matter of good fortune, and in general edgewise views are not very helpful.

The effects would be more pronounced with denser metals, with capacitors of lower X-ray stopping power, or with capacitors having a more even distribution of X-ray stopping power, and would be less pronounced with the opposites of these conditions. The amount of silver electrode in the above capacitor was more than would be present in the average case, but the silvering was fairly even in thickness. In some types of silvered mica the silvering is less even, giving a mottled effect in a radiograph and making detection of particles more difficult.

(2.3.3) Observing Mechanical Damage due to Particles of Known Size.

Stacks of unsilvered mica plates 0·001 5 in thick were assembled with measured particles inserted in them, and were compressed in the same way as in normal manufacture. The pressure was then released and the stack was opened and examined. Impressions were found on the adjacent plates and for a number of plates on either side, depending on the size of the particle. Detailed examination of the damaged plates showed that:

(*a*) The shape of the particle is often copied in the nearest marks.
(*b*) Clear impressions show on two plates and with decreasing intensity successively on others on either side in accurate location.
(*c*) There are cracks round the impressions, but fewer and shorter than in the case of excess voltage impressions.
(*d*) The location of the point of failure may obviously be anywhere on a plate.

Table 3

SOME CHARACTERISTICS OF MECHANICAL DAMAGE DUE TO FOREIGN PARTICLES

Particle	Diameter	Number of micas in stack	Number marked	Remarks
	in			
Brass wire . .	0·030	40	20 + 20	whole stack marked
Brass sphere	0·023	40	8 + 8	
Brass sphere	0·015	40	4 + 4	
Brass sphere	0·007	40	1 + 1	
Sand . .	0·015	40	1 + 1	grain of sand broken
Grit . .	0·003	40	1 + 2	

In an investigation on causes of failure of drystack mica capacitors in 1944, Thomas[4] estimated that of those examined about 50% were caused by mechanical damage from dust particles introduced during assembly.

(2.4) Faulty Mica Laminations

Mica is formed naturally in the earth's crust during the cooling of molten rock, and hence is liable to contain adventitious faults.

There is no British Standard covering visual quality of mica laminations, but A.S.T.M. D351–57T[5] adopts a perfectly clear transparent flat specimen of mica as a visual standard of perfection, independent of the basic colour of the mica. The highest quality, V–1 clear, is hard and does not contain any crystallographic discolorations, air inclusions, cloudy stains, black, red or green dots or stains, clay stains, waviness, sandblast, stones and holes, buckles, reeves, ridges, tears, cracks, hairline cracks, wedges, tangle sheets or herring bones.

A decreasing standard is represented by grades V–2, V–3, etc., and the lowest, V–10A, permits the above faults with the exception of the last ten.

A.S.T.M. D748–52T[6] discards the visual classification of mica films (for capacitors) and does not discriminate against colour, spots or stains, provided that the mica meets specific electrical and physical requirements, namely electrical conductivity (using a spark-coil test set with which conducting spots, veins, or areas are shown up by sparking or a glow, and pinholes, tears, cracks, etc., by breakdown), Q-factor at 1 Mc/s, electric strength, weight loss on heating to 600° C, thickness uniformity, and visual qualities.

For capacitors used in the transatlantic telephone cable it was required that mica laminations should be uniform in shape and thickness, hard, and free from inclusions, stains, cracks, creases waves and buckles, but a proportion of iridescence and air inclusions were permitted. An investigation in the Post Office Research Branch in 1954 showed that iridescence had very little effect on the quality of mica for capacitor purposes.

Thomas[4] has described how minute black spot in mica plates can be the cause of subsequent failure, and how a number of capacitors made up from stained mica plates showed no failures. In a further report[7] he stated that he was unable to recommend visual inspection for the purpose of reducing the percentage of defective finished mica capacitors.

When failure is due to a defect in the mica it often takes place at a low voltage, and as the energy released is small there are no cracks. Also, adjacent plates are unmarked unless an exceptionally large amount of energy is released at breakdown, in which case the puncture will be abnormally large.

(2.5) Mica Plates Damaged during Manufacture of Capacitors

In an investigation on causes of failure of mica capacitors in 1944, Thomas[4] ascribed about one-third of the cases examined to prior mechanical damage of the plates, such as abrasion of the surface and knocks, and pointed out that previous mechanical damage of this kind will usually be visible adjacent to the puncture. Several illustrations are attached to the report. Surface scratches were found not to reduce the electric strength of the mica.

In a study made in the Post Office in 1954–55 of the variability under working conditions of silvered-mica capacitors in certain amplifiers, it was found that an appreciable proportion of the micas had been damaged mechanically during manufacture and it was considered that the failure, which took place at comparatively low voltages and with a high-value series resistor, were due to an ionic mechanism acting through these or inherent faults in the micas. Scratches with a depth as small as 0·00001 in can be seen plainly with good illumination but are not thought to have a harmful effect, but obviously deep scratches weaken the mica electrically and should not be allowed.

Some recent failures of mica capacitors have been found to be due to the burr on the edge of an enveloping metal clamp cutting nearly through the mica cover-plate.

(2.6) Faulty Electrodes

A few types of defect which are normally eradicated in the manufacturer's inspection may be classed as due to faults in the electrodes. On rare occasions a mica capacitor is found with a loose electrode which can be detected by a high value of tan δ and by the capacitor emitting a faint note when tested at audio frequency. Generally, of course, the capacitor will be clamped so tightly that the effect is very slight.

High values of tan δ may be due to carbonaceous matter from the organic binder used in the silver printing ink not being completely burnt off in the firing process. Alternatively, they may be due to over-firing of the lamination to the extent that the silver has attacked the mica.

High values of tan δ at 1 Mc/s in capacitors having normal values at 1 kc/s may be due to electrodes which are too thin and give a series resistance which is significant at the higher frequency only.

(2.7) Faulty Connections

Where solder is used to make the connection to a silver electrode, the thermal contraction of the solder on cooling has sometimes been found to cause it to tear away from the silver, but to leave so narrow a gap that it is invisible to the naked eye.

Where pressure connections are made to the electrodes by tabs of thin metal ribbon (virtually forming intentional dry joints), it has been found that, unless substantial contact pressure is present, a surface oxide film may form on it in a long period of time if it is of copper or aluminium, and a sulphide film if it is of silver. If this happens, tan δ increases with the passage of time until the electrode is virtually disconnected. The film may sometimes be broken down by the sudden application of a potential.

(3) CAUSES OF IMPERFECT PERFORMANCE

(3.1) Silver-Ion Migration

Reference has been made in Section 2.5 to work done in 1954 when it was suspected that an ionic mechanism acting through faults in the plates was the cause of failures in silvered-mica capacitors. At this time silver-ion migration through phenolic laminates was well known, and so work was started to see if similar phenomena could be made to occur on or in mica.

A number of silvered-mica laminations were carefully removed from some capacitors and cleaned by washing successively with trichloroethylene and absolute alcohol. They were then placed inside a vessel humidified by a beaker of distilled water and were wired up in series with individual microammeters, 1-megohm resistors, and a 240-volt battery. The two portions of silver at different polarities were separated by a clear strip of mica 2–3 mm wide. Similar samples were set up in the laboratory atmosphere but protected from contamination.

After three months, some of the humidified samples were passing a small current and were removed for microscopic examination. A large number of dark markings had appeared on the clear mica in the gap between the two silvered areas. They appeared to stretch out from the negative area and are shown in Fig. 4. As it was obvious that the silver-ion migration had reached an advanced stage, the experiment was repeated

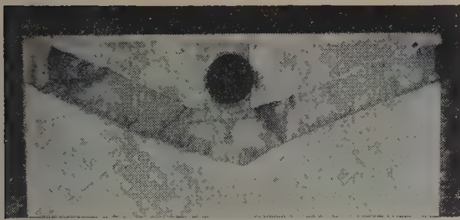


Fig. 4.—Silvered mica showing silver-ion migration.
Magnification, ×3.

and the samples were examined each day. After a week, markings were showing clearly, but faintly compared with the previous test. No traces of silver-ion migration were detected in the samples that had been at room humidity. The humidified samples can be considered to have been at 95–100% relative humidity, and the room samples at 60–70% relative humidity.

Subsequently, two papers[8,9] on the subject were published suggesting that the effect is based on the anodic solution of silver oxide in water,

$$Ag_2O + H_2O \leftrightarrows 2AgOH \leftrightarrows 2Ag^+ + 2OH^-$$

and that the only features necessary to produce silver-ion migration are silver electrodes with a difference of potential, a surface (as with mica) or internal surfaces (as with cellulose) on which water can be adsorbed, and a humidity high enough for this to take place to a significant extent. Under some circumstances silver migration has been observed at humidities as low as 75% relative humidity in the presence of traces of hygroscopic salts.

The studies described refer only to surface migration, which is fairly easily observed. The effect never takes place through a sound mica lamination, but it seems probable that it can take place through a flaw which is equivalent to a hole or porosity in the mica. This is almost impossible to prove by dissection of faults in silvered-mica capacitors, but it could be decided by silvering a number of micas having known faults of this type (leaving a very large surface clearance, which could also be waxed to prevent surface moisture adsorption) and observing their behaviour under high humidity and a potential.

(3.2) Growth of Whiskers

When undisturbed for periods of some months it is possible for fine metal whiskers[11] to grow out from the surface of certain metals, in particular from solid or plated tin, cadmium, or zinc, and especially at angular sites where the metal beneath the plating has been heavily stressed. Silver plating on nickel silver is known to grow whiskers in the presence of atmospheres containing a trace of sulphur dioxide, but the composition of the whiskers is unknown. Black whiskers (presumably sulphide) will also grow on silver (and copper) but only in the presence of sulphur. They are usually shorter and rougher in appearance.

The main relevance of this to mica capacitors is the possibility of the growth of whiskers from a lead or connection to another part of the circuit in close proximity.

(3.3) Islands in Silver Electrode causing Capacitance Fluctuation

This effect has been investigated thoroughly by the Electrical Research Association[10] and only a description and the explanation will be given here.

Certain types of silvered-mica and silvered-ceramic capacitors are liable to spontaneous capacitance fluctuations when under r.f. stress. This shows clearly if an a.f. note is obtained by beating two r.f. oscillators, one of which contains the faulty capacitor in its tuned circuit, when there will be periodic jumps in the frequency of the beat note corresponding to capacitance changes of the order of 0·01–0·1 pF. A rapid flutter sometimes arises from changes of about 0·001 pF. These fluctuations are caused by areas of silver separated from the main electrode by a high-resistance bridge of thinly-deposited metal of varying resistance. It is occasionally possible to identify the offending area visually, but proof rests mainly on the fact that, when the electrodes of such capacitors are backed with extra metal by electroplating, the fault disappears.

(3.4) Faults arising from Impregnating, Filling and Coating Compounds

Mica-capacitor stacks are often impregnated with oil or wax to improve insulation resistance, raise the a.c. ionization voltage and retard the ingress of moisture. Unsuitable impregnants used in this way may lead to environmental cracking of cast- or moulded-resin housings. Draining of the impregnant in unsuitable designs may result in a gradual drift of capacitance.

Filling compounds in older types of mica capacitors (with metal-foil electrodes) were sometimes bituminous materials with a high value of tan δ and relatively low insulation resistance, and these in impregnated or unimpregnated types were liable to migrate or diffuse between the plates, causing a little drift in capacitance but particularly a fall in insulation resistance and a rise in tan δ.

Some form of material to surround and protect the capacitor has been used since mica capacitors were invented. Occasionally the following faults have been met: ebonite surrounds lead to sulphiding of the electrodes and rise in tan δ; wax, while satisfactory in dry situations, may crack and allow ingress of moisture; plasticizers or other liquids from moulded or cast surrounds may creep or diffuse between the plates, alter the capacitance slightly, and cause an increase in tan δ.

(3.5) Faults arising from Moisture

Reference has been made in Section 3.1 to the role of water in the migration of silver ions, and in Section 3.3 to ingress of water through cracks in wax. It should be added that no organic coating material acts as a complete barrier to moisture-vapour diffusion, and mica in such coatings will gradually absorb[12] an amount of moisture corresponding to the average humidity of the surrounding atmosphere over a long period and remain so with small fluctuations on either side of this value. This, however, is a valuable protection against temporary peak periods of very high humidity. The protection is not obtained if there are cracks or fissures in the coating, e.g. at the points where the leads enter.

If moisture is taken up, either by foil-mica or silvered-mica capacitors, there is a slight increase[12] (of the order of 0·1% from 0–60% relative humidity) in capacitance, a marked increase[12] (of the order of 0·0001 to 0·0006 from 0–60% relative humidity) in tan δ and a decrease in insulation resistance.

(3.6) Intermittent Faults

Intermittent faults are sometimes the most difficult to locate and diagnose. They generally fall into two broad classes:

(*a*) Faults usually arising under low-voltage conditions which clear themselves abruptly when the voltage (or available energy) is raised.

(*b*) Faults which come on and off in phase with some operating condition, e.g. temperature, humidity, voltage, or mechanical pressure or tension.

Disconnections of class (*a*) are nearly always due to a dry joint at which an oxide film slowly builds up but which can be

Table 4

TABULATION OF EFFECTS OBSERVED

Class of fault	Observed effects	Possible mechanism
Breakdown	Short-circuited or low insulation resistance. Puncture on edge of electrode. No signs of foreign particles or marks on micas other than those described in Section 2.1	Excessive voltage. (Perhaps capacitor unimpregnated or oil-impregnated)
	High insulation resistance on low voltage but will only withstand low voltage. Otherwise as last	Excessive voltage. (Perhaps wax-impregnated)
	Either of above with puncture not at edge of electrode, and particles or marks as described in Sections 2.3, 2.4, 2.5	Foreign particles. Faults in mica or damage to mica
	High insulation resistance and will withstand 1 000 volts d.c. for 1 min. No puncture but marks as in Section 2.2	Flashover
Low insulation resistance	Low insulation resistance but not broken down. Marks as in Section 2.2	Flashover
	Low insulation resistance but clears on raising voltage Marks as in Section 2.2	Flashover
	As last, but marks as in Section 3.1	Silver-ion migration
	As last, but no marks. Perhaps evidence of whiskers. Clears suddenly under voltage	Metal whiskers
	Insulation resistance improves greatly on drying	Moisture film between terminals. Moisture absorbed by housing material Moisture in mica. Silver-ion migration
Capacitance change	Capacitance reduced to a few picofarads	Disconnection in lead or between lead and electrode
	Capacitance reduced by value of one plate	Disconnection of one plate
	Capacitance reduced by less than value of one plate	Disconnection of adjusting plate or part of one electrode
	Capacitance changed 0·2–0·5% with little or no change in tan δ	Mechanical drift of plates with time. Loss of impregnant. Compression or relaxation of housing
	Capacitance occasionally fluctuates a fraction of 1 pF under r.f. stress. No change in tan δ	Islands in silvered electrode
Rise in tan δ	Capacitance increases a few per cent and tan δ increases	Lossy impregnating, filling or coating compound migrating between plates
	Tan δ increases but no change in capacitance	Dry joint in lead gradually oxidizing
Noise	Capacitance fluctuates widely on handling	Loose lead or electrode
	No capacitance change	Dry joint
	Low insulation resistance which may clear	Silver-ion migration or metal whiskers
	Noise occurs frequently and regularly with nominal working alternating voltage but only initially with direct voltage	Gaseous ionization in spaces between or in micas

broken down by increased voltage: short-circuits or low insulation resistance faults of class (*a*) are sometimes due to metal whiskers or silver-ion migration, but may also be singly-occurring results of types described in Sections 2.1–2.5.

With class (*b*) intermittent faults, diagnosis is greatly helped if they can be related to an operating condition, and frequent measurements made of capacitance and tan δ at a low voltage while the capacitor is successively warmed, cooled, pressed or pulled.

(4) TABULATION OF EFFECTS OBSERVED

Table 4 offers suggestions regarding possible mechanisms of failure of mica capacitors when only observations by the investigator are available. Any accurate detailed information about the history of the capacitor and conditions at the time of failure will greatly increase the probability of a quick and accurate diagnosis. Of necessity, the indications in the Table are vague and reference should be made to the main text when the possibilities have been narrowed down.

(5) DISCUSSION

There are many possible mechanisms of failure of mica capacitors, some inherent in the mica but most resulting from faulty manufacture. Few of them can be detected easily by inspection after manufacture. The nature of a fault can often (though not always) be deduced by thorough inspection, including dissection, of a faulty capacitor after failure; it is often helpful in doing so to know the working conditions, both electrical and ambient, and particularly the exact conditions at the time of failure. There are, however, many types of failure where the conditions at the time of failure can be deduced from the dissection and detailed examination of the capacitor.

It is suggested that, to use the information in the paper to the best advantage in any particular case, Table 4 should be consulted first, followed by reference to the body of the paper and the illustrations, and lastly, if more detail is required, to the references.

This work does not imply that the proportion of mica capacitors which fail in service is excessive. Some of the causes of failure are liable to occur in other types of capacitor as much as in those using mica.

(6) ACKNOWLEDGMENTS

Acknowledgment is made to the Engineer-in-Chief of the Post Office and to the Controller of H.M. Stationery Office for permission to publish the paper.

The author is indebted to his colleagues at the Post Office Research Station for the help they have given, and in particular to Mr. C. R. Schroder for carrying out many of the dissections and measurements.

(7) REFERENCES

(1) SCHUMANN, W. O.: 'Elektrische Durchbruchfeldstärke von Gasen' (Julius Springer, 1923), p. 25.
(2) MAXWELL, F. A., and BENEDICT, R. A.: 'Theory of Gaseous Conduction and Electronics' (McGraw Hill, 1941), p. 297.
(3) WHITEHEAD, S.: 'Dielectric Breakdown of Solids' (Oxford University Press, 1951), p. 223.
(4) THOMAS, A. M.: 'Defects in Mica Plates and Capacitors and the Use of Emergency Substitute Grades', E.R.A. Report Ref. D/T30, 1944.
(5) A.S.T.M. D351–57T: 'Natural Muscovite Mica based on Visual Quality.' First issued, 1949.
(6) A.S.T.M. D748–52T: 'Natural Block Mica and Mica Films Suitable for Use in Fixed Mica-Dielectric Capacitors.' First issued, 1952.
(7) THOMAS, A. M.: 'The Elimination of Defective Mica Plates for Capacitors by Visual Inspection', E.R.A. Report Ref. D/T34, 1946.
(8) KOHMAN, G. T., HERMANCE, H. W., and DOWNES, G. H.: 'Silver Migration in Electrical Insulation', *Bell System Technical Journal*, 1955, **34**, p. 1115.
(9) WILLIAMS, J. C., and HERRMAN, D. B.: 'Surface Resistivity of Non-porous Ceramic and Organic Insulating Materials at High Humidity, with Observations of Associated Silver Migration', *Transactions of the Institute of Radio Engineers*, February, 1956, **RQC–6**, p. 11.
(10) CHURCH, H. F.: 'Spontaneous Capacitance Fluctuations in Silvered Ceramic and Silvered Mica Capacitors', E.R.A. Report Ref. L/T181, 1947.
(11) COMPTON, K. G., MENDIZZA, A., and ARNOLD, S. M.: 'Filamentary Growths on Metal Surfaces—"Whiskers"', *Corrosion*, 1951, **7**, p. 327.
(12) RAYNER, G. H., and FORD, L. H.: 'The Performance of Dried and Sealed Mica Capacitors', *Journal of Scientific Instruments*', 1954, **31**, p. 3.

(8) APPENDIX: SOME METHODS OF DISSECTION OF MICA CAPACITORS

(8.1) Removal of Waxes and Lacquers

Most of the waxes used for coating capacitors are hydrocarbon waxes with melting points in the range 70–80° C, and can be removed by heat or by solvents with very little disturbance of the capacitor unit as follows:

(*a*) Suspend the waxed capacitor over a small beaker in an oven at 80–90° C for an hour or so, or

(*b*) Immerse it in a boiling aromatic hydrocarbon such as benzene (b.p. 80° C) or toluene (b.p. 111° C)*.

Chlorinated hydrocarbons will dissolve the waxes more quickly but are liable to attack any aluminium that may be present.

Cellulose lacquers are sometimes used for coating capacitors and can be removed by immersion in cold or hot acetone (b.p. 56° C) or methyl ethyl ketone (b.p. 80° C).*

(8.2) Removal of Phenolformaldehyde and Other Moulded Casings

Phenolformaldehyde and similar mouldings can often be removed by cutting off the leads and grinding away the edges of the mouldings until the edges of the micas are just exposed, when the remainder of the moulding can be gently prized off. This, however, requires care and uses up a great deal of an operator's time if many capacitors are involved.

Another method is to place the cased capacitor in 16% aqueous caustic soda solution at about 60° C for about 5 hours (small capacitors) to 30 hours (large capacitors), when the structure of the casing will be broken down to an extent depending on the individual resin and degree of cure. The casing and unit should be washed in water and finally allowed to stand for some hours in several changes of distilled water before draining and drying.

Certain materials, e.g. aluminium or zinc, are attacked by caustic soda, but provided that such materials are absent a clean removal of the casing can often be made without damage to, or destroying evidence in, the capacitor unit.

Occasionally one meets capacitors having moulded casings of polystyrene, acrylic resins, polythene, p.t.f.e. or p.c.t.f.e. The first three of these can generally be melted off at temperatures of the order of 120–150° C, but if it is not desired to heat them so strongly as this, solvents may be used. Polystyrene casings can usually be dissolved in aromatic hydrocarbons, such as benzene, toluene or xylene.* Acrylic resins will generally dissolve in ketones such as acetone or methyl ethyl ketone.* Polythene casings can often be made to crack and partially granulate by immersion in acetone, to such an extent that they can be removed.* P.T.F.E. and p.c.t.f.e. are so resistant to heat and solvents that only mechanical methods are likely to be successful.

(8.3) Removal of Potting Compounds

Many capacitors are encapsulated in epoxy-type resins. At the commencement of this work information on the removal of such resins was limited to mechanical or thermal methods, neither of which could be employed without almost certain damage to the capacitor unit. Chemical means would be more suitable, but information was scarce as to the solubility of epoxy resins. Sample blocks of the resin were cast and the effect of different classes of solvent upon them was studied, with the results shown in Table 5.

Methyl ethyl ketone was obviously the solvent to concentrate upon, and so a capacitor was placed in this liquid in a stoppered container, at 50° C. After 20 hours most of the epoxy resin had crumbled away from the unit, and standing thus, with periodic agitation, all the resin was removed without damage to the capacitor unit. This was successfully repeated many times with other capacitors. The work was carried out on one type of epoxy resin only, but others would be expected to behave in a similar manner, though they may be slightly more or less resistant to the solvent.

Subsequently, the author has been informed that one manufacturer uses boiling methylene dichloride (b.p. 40° C) for this purpose.

(8.4) Separation of Silvered Micas in a Stack

With some makes of silvered-mica capacitor it is a simple mechanical operation to separate out the individual silvered micas, after proceeding as described in Sections 8.1, 8.2 and 8.3; with others a little unsoldering is involved, but with at least one make the adjacent silver surfaces are fused together, and it is impossible to separate the laminations as silvered micas without damage. They may be separated as micas and double-thickness silvers by warming in a 16% caustic soda solution for a few minutes, or by boiling in a 10% solution of sodium carbonate for several hours. The micas can, however, be recovered intact and unaltered without their silver by the method of Section 8.5.

The effect of this treatment on the mica is negligibly small, but it may remove part of the characteristic appearance of tracking or flashover.

* Suitable precautions should be taken against fire hazards.

Table 5

EFFECTS OF SOLVENTS ON EPOXY RESINS

Class	Individual	Boiled, 3–5 min	Cold, 17 h
Water	Water	No effect	No effect
Alcohols	Ethyl alcohol	Little effect	Slightly softened
Esters	Ethyl acetate	Little effect	No effect
Ketones	Methyl ethyl ketone	Little effect	Crumbled
Chlorohydrocarbons	Trichlorethylene	Little effect	No effect
Carboxy acids	Acetic acid	Slightly softened	Softened
Polyhydroxy alcohols	Glycerol	Slightly softened	Softened
High b.p. amines	Aniline	Slightly softened	Softened
High b.p. phenols	Phenol	Slightly softened	Softened

(8.5) Removal of Silver from Silvered Micas

Silver electrodes can be removed from micas without any attack on the mica by heating gently in 25% (approximately) nitric acid. Solution takes place in a few minutes with single silvered-mica laminations, but with stacks, attack can only take place along the edges and 2–8 hours must be allowed. Finally, the micas should be washed with numerous changes of distilled water and dried on filter paper.

The correct sequence and orientation of micas in a stack can be retained by lightly binding it with platinum wire before treatment.

21.396.962.25

The Institution of Electrical Engineers
Paper No. 3264 E
July 1960

ANALYSIS OF A FREQUENCY-MODULATED CONTINUOUS-WAVE RANGING SYSTEM

By A. J. HYMANS, M.Sc., A.Inst.P., Graduate, and J. LAIT, M.A.

(*The paper was first received* 4th *September*, 1959, *and in revised form* 4th *February*, 1960.)

SUMMARY

Some aspects of an f.m. c.w. radar with a sawtooth frequency sweep are considered. The exact beat note for a discrete target is calculated and its Fourier transform is obtained. A scheme previously given by Gnanalingam for producing a coherent system is shown to be only approximately valid, and an alternative method is proposed. The effect of Doppler shift on the return is discussed. Range discrimination is examined critically.

LIST OF SYMBOLS

c = Velocity of propagation, m/s.
$F(\omega)$ = Complex Fourier transform.
$F^*(\omega)$ = Complex conjugate of $F(\omega)$.
F_1, F_2, F_3, F_4 = Individual terms in Fourier transform of beat note.
f_1, f_2 = First and second intermediate frequencies, c/s.
G = Numerical constant.
k = Order of spectral line.
m = Order of zero in envelope of beat-note spectrum.
n = Serial number of sweep interval.
P = Echo power, watts.
P_A, P_B = Echo powers received from targets at ranges R_A, R_B, watts.
R = Target range, m.
R_0 = Target range at zero time, m.
R_A, R_B = Ranges of targets producing complementary beat notes, m.
T_s = Sweep duration time, sec.
t = Time variable, sec.
t_n = Time variable measured from mid-point of nth sweep interval, sec.
V_e, V_g = Amplitudes of echo and ground-wave voltages, volts.
V_s, V_{s1}, V_{s2} = Detector outputs, volts.
v = Radial component of target velocity, m/s.
2α = Angular-frequency sweep rate, rad/s^2.
$\delta(\omega)$ = Dirac delta function.
τ = Time delay of echo, sec.
ϕ_e = Instantaneous phase of echo, rad.
ϕ_g = Instantaneous phase of ground wave, rad.
ϕ_i = General instantaneous phase, rad.
ω = Angular frequency, rad/s.
ω_0 = Undeviated carrier angular frequency, rad/s.
ω_{B1}, ω_{B2} = Lower and upper maxima of envelope of beat-note spectrum, rad/s.
ω_i = Instantaneous angular frequency, rad/s.

(1) INTRODUCTION

The application of f.m. techniques to radar ranging has been discussed by Keep,[2] Tucker,[3] Kay[4] and others. The advantages and disadvantages compared with pulse radars have been set out at length by them. In particular, it has been shown that maximum sensitivity (and hence range) depends on mean power, and that the f.m. system has the advantage that it avoids the necessity for the high peak powers in the pulse system. The f.m. system suffers, however, from the indirect method of obtaining range information, and the receiver becomes complicated and expensive if a simultaneous presentation analogous to the A-scan is required. The question of mean power required is determined by the noise in the system, and the possibility of using a coherent system for frequency modulation was pointed out by Gnanalingam,[1] who produced a vertical-incidence ionosphere sounder of great sensitivity. In all cases of practical application to date, the maximum range of targets of interest has been quite short (e.g. airborne altimeters). By this, we mean that the time of travel of the signal is small compared with the repetition time of the sweep. Gnanalingam based his analysis on an approximation which was valid only in the limit of zero range. In this paper, the more general case is examined, and a complete expression for the beat note and its spectrum is obtained.

The present paper, which arose from design studies for an ionospheric sounder, re-examines the mathematical basis of f.m. ranging and goes on to propose a ranging system which avoids some of the ambiguities of earlier systems, but retains coherence in the method of detection. The equipment concerned is in the process of construction and results of measurements will be made available in due course.

In the analysis, considerable attention is given to the exact mathematical formulation of the functions concerned, so that misleading approximations may be avoided.

(2) PRINCIPLE OF THE SYSTEM

In the system under investigation, the transmitter produces a c.w signal of constant amplitude, whose frequency is varied in sawtooth fashion (see Fig. 1).

The receiver picks up some of the power from the transmitter (a short distance away) and also the echo signal after delay of time τ due to its travel to and from the target. The direct signal from the transmitter is known as the ground wave. The two oscillations are made to beat together in a non-linear device (e.g. a diode detector) and the beat note is found to contain two distinct tones (see Fig. 1).

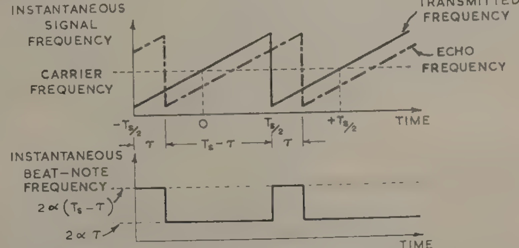


Fig. 1.—Production of the beat note.

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.
Mr. Hymans and Mr. Lait are at the Royal Military College of Science.

This beat note is the equivalent of the video-frequency pulse in a pulse radar and contains all the information. Since the beat note is a repetitive waveform with repetition time T_s its Fourier transform must consist of a spectrum of lines spaced at intervals $\omega_s = 2\pi/T_s$.

In the analysis that follows two approximations are made for simplicity. They are as follows:

(a) The frequency increases linearly with time and has instantaneous flyback, although this is not, of course, possible in practice. With the type of radar considered, however, the assumption of negligible flyback time is thought to be a reasonable approach to the truth.

(b) The exact phase of both signals is calculated, taking into account the fact that the echo is merely the direct signal delayed by a time τ. The medium through which the signal travels is assumed to be non-dispersive.

(3) MATHEMATICAL FORMULATION OF THE SIGNAL

The instantaneous frequency, ω_i, is given by the following set of expressions:

$$\begin{aligned}\omega_i &= \omega_0 + 2\alpha t, \text{ where } -\tfrac{1}{2}T_s < t < +\tfrac{1}{2}T_s \\ &= \omega_0 + 2\alpha(t - T_s), \text{ where } \tfrac{1}{2}T_s < t < +\tfrac{3}{2}T_s \\ &\quad \cdots \\ &= \omega_0 + 2\alpha(t - nT_s), \text{ where } \tfrac{1}{2}(2n-1)T_s < t < \tfrac{1}{2}(2n+1)T_s \qquad (1)\end{aligned}$$

For convenience in handling the expressions, the sweep rate has been taken as 2α rather than α, and the origin of time is set at the centre of one sweep cycle.

It has been found useful to make the substitution

$$t_n = t - nT_s \qquad (2)$$

so that the expressions can be generalized to give, in the nth interval,

$$\omega_i = \omega_0 + 2\alpha t_n; \ -\tfrac{1}{2}T_s < t_n < +\tfrac{1}{2}T_s \qquad (3)$$

Instantaneous frequency is not a physically measurable variable; only voltages and currents can be so regarded. The voltage in the transmitter is proportional to the sine or cosine of the phase angle ϕ_i, where

$$\phi_i = \int^t \omega_i dt + \text{constant} \qquad (4)$$

Thus we must obtain this integral of the instantaneous frequency by substitution from eqn. (3) into eqn. (4). The constant of integration can be evaluated by arbitrarily putting $\phi_i = 0$ at $t = 0$ (there is no loss of generality here) and then stipulating that phase shall be a continuous function. Such a restriction is necessarily imposed by the continuity of currents and voltages in real circuits.

Taking the interval of zero order as an example, we then have, from eqn. (1),

$$\omega_i = \omega_0 + 2\alpha t_0$$

and

$$\phi_i = \int_0^{t_0} (\omega_0 + 2\alpha t_0) dt_0$$

$$= \omega_0 t_0 + \alpha t_0^2 \qquad (5)$$

Similarly, the general expression will be found to be

$$\phi_i = \omega_0 t_n + \alpha t_n^2 + n\omega_0 T_s \qquad (6)$$

An examination of eqn. (6) shows that ϕ_i is continuous at the ends of each sweep.

(4) PRODUCTION OF THE BEAT NOTE

Let the reference or ground wave, which passes direct from transmitter to receiver, induce an oscillatory voltage in the aerial $V_g \sin \phi_g$, where ϕ_g is the expression in eqn. (6), with phase arbitrarily zero at $t = 0$. The echo from a stationary reflecting target will induce a voltage $V_e \sin \phi_e$, where ϕ_e is a function similar to ϕ_g but delayed in time by τ, given by

$$\tau = 2 \times \frac{\text{Range of target}}{\text{Velocity of propagation}} = \frac{2R}{c} \qquad (7)$$

In most cases we may assume that $V_g \gg V_e$. These two oscillations are 'mixed', 'heterodyned', or made to 'beat' together, in some non-linear device. The resulting signal or beat note will contain a d.c. (zero-frequency) term and a product term $GV_eV_g \sin \phi_e \sin \phi_g$, where G is a numerical constant, and other higher-order products.

In general, only the lowest-order product will have a significant amplitude, apart from the large d.c. term.

The product may be expanded as a difference, namely

$$\tfrac{1}{2}GV_eV_g[\cos(\phi_g - \phi_e) - \cos(\phi_g + \phi_e)]$$

The phase-sum term is an oscillation at radio frequency and is removed by filtering. The difference term is the desired oscillation at video frequency, and contains all the range information. We are thus interested in the function $\tfrac{1}{2}GV_eV_g \cos(\phi_g - \phi_e)$.

It will be seen that $\phi_g - \phi_e$ has two forms in the nth interval. They are

(i) During time $-\tfrac{1}{2}T_s < t_n < -\tfrac{1}{2}T_s + \tau$,

when $\phi_e = \omega_0(t_{n-1} - \tau) + \alpha(t_{n-1} - \tau)^2 + (n-1)\omega_0 T_s$

and $\phi_g = \omega_0 t_n + \alpha t_n^2 + n\omega_0 T_s$

$$\phi_g - \phi_e = \omega_0\tau - \alpha(T_s - \tau)^2 + 2\alpha(\tau - T_s)t_n \qquad (8)$$

(ii) During time $-\tfrac{1}{2}T_s + \tau < t_n < +\tfrac{1}{2}T_s$,

when $\phi_e = \omega_0(t_n - \tau) + \alpha(t_n - \tau)^2 + n\omega_0 T_s$

and ϕ_g is as in (i),

$$\phi_g - \phi_e = \omega_0\tau - \alpha\tau^2 + 2\alpha\tau t_n \qquad (9)$$

(5) FREQUENCY ANALYSIS OF THE BEAT SIGNAL

The oscillation $\tfrac{1}{2}GV_eV_g \cos(\phi_g - \phi_e)$ is the beat note between the echo and the reference wave. This oscillation is not a pure sinusoid and can be analysed into separate harmonic components by the use of the Fourier transform. If ω is the general variable in the transform, we define the Fourier transform as

$$\mathrm{F}(\omega) = \int_{-\infty}^{+\infty} \tfrac{1}{2}GV_eV_g \cos(\phi_g - \phi_e)\varepsilon^{-j\omega t} dt \qquad (10)$$

This integral may be reduced to

$$\mathrm{F}(\omega) = \tfrac{1}{2}GV_eV_g \sum_{-\infty}^{+\infty} \int_{\frac{1}{2}(2n-1)T_s}^{\frac{1}{2}(2n+1)T_s} \cos(\phi_g - \phi_e)\varepsilon^{-j\omega t} dt \qquad (11)$$

or by use of the substitution $t_n = t - nT_s$,

$$\begin{aligned}\mathrm{F}(\omega) = \tfrac{1}{2}GV_eV_g \sum_{-\infty}^{+\infty} \varepsilon^{-jn\omega T_s} \Bigg\{ &\int_{-\frac{1}{2}T_s}^{-\frac{1}{2}T_s+\tau} \cos[\omega_0\tau - \alpha(T_s - \tau)^2 + 2\alpha(\tau - T_s)t_n]\varepsilon^{-j\omega t_n} dt_n \\ &+ \int_{-\frac{1}{2}T_s+\tau}^{+\frac{1}{2}T_s} \cos(\omega_0\tau - \alpha\tau^2 + 2\alpha\tau t_n)\varepsilon^{-j\omega t_n} dt_n \Bigg\} \qquad (12)\end{aligned}$$

The first integral in the outer bracket gives two terms:

$$F_1(\omega) = \frac{\tau}{2}\,\frac{\sin\,[\omega - 2\alpha(T_s - \tau)]\frac{\tau}{2}}{[\omega - 2\alpha(T_s - \tau)]\frac{\tau}{2}}\exp - j[\omega_0\tau - \tfrac{1}{2}\omega(T_s - \tau)] \quad . \quad . \quad . \quad . \quad (13)$$

and

$$F_2(\omega) = \frac{\tau}{2}\,\frac{\sin\,[\omega + 2\alpha(T_s - \tau)]\frac{\tau}{2}}{[\omega + 2\alpha(T_s - \tau)]\frac{\tau}{2}}\exp j[\omega_0\tau + \tfrac{1}{2}\omega(T_s - \tau)] \quad . \quad . \quad . \quad . \quad (14)$$

The second integral in the outer bracket gives two terms:

$$F_3(\omega) = \frac{T_s - \tau}{2}\,\frac{\sin\,(\omega - 2\alpha\tau)\left(\frac{T_s - \tau}{2}\right)}{(\omega - 2\alpha\tau)\left(\frac{T_s - \tau}{2}\right)}\exp j(\omega_0\tau - \tfrac{1}{2}\omega\tau) \quad . \quad . \quad . \quad . \quad (15)$$

and

$$F_4(\omega) = \frac{T_s - \tau}{2}\,\frac{\sin\,(\omega + 2\alpha\tau)\left(\frac{T_s - \tau}{2}\right)}{(\omega + 2\alpha\tau)\left(\frac{T_s - \tau}{2}\right)}\exp - j(\omega_0\tau + \tfrac{1}{2}\omega\tau) \quad . \quad . \quad . \quad . \quad (16)$$

The remaining factor in $F(\omega)$ is a delta function,

$$\text{for} \quad \sum_{-\infty}^{+\infty} \varepsilon^{-jn\omega T_s} = \omega_s\delta(\omega - k\omega_s),\ k = 0, \pm 1, \pm 2 \ldots \quad (17)$$
$$\text{where } \omega_s = 2\pi/T_s$$

Thus the full expression for the Fourier transform of the beat note is

$$F(\omega) = \tfrac{1}{2}GV_eV_g\omega_s \sum_k \delta(\omega - k\omega_s)$$
$$\times\,[F_1(k\omega_s) + F_2(k\omega_s) + F_3(k\omega_s) + F_4(k\omega_s)] \quad (18)$$

(6) EXAMINATION OF THE SPECTRUM

(6.1) General Nature of the Solution

From an inspection of eqns. (13)–(16) and (18) it will be seen that

$$F(\omega) = F^*(-\omega)$$

where $F^*(\omega)$ is the complex conjugate of $F(\omega)$. This is to be expected since we have here taken the Fourier transform of a real function. The general form of the spectrum is a set of lines spaced along the ω scale at interval ω_s.

The function shows four distinct maxima, at $\omega = \pm\,2\alpha\tau$ and at $\omega = \pm\,2\alpha(T_s - \tau)$.

The variable ω is, however, merely a mathematical tool which permits one to examine the properties of real electrical functions. To obtain the phases and amplitudes of these real components one must combine the positive- and negative-frequency terms in pairs, when it will be seen that the real parts of a pair of complementary terms (such as F_1 and F_2) reinforce and their imaginary parts cancel, as pointed out by Woodward.[5]

From the terms $F_1(\omega)$ and $F_2(\omega)$ of eqns. (13) and (14) there arises a single real harmonic component

$$\tau\,\frac{\sin\,[k\omega_s - 2\alpha(T_s - \tau)]\frac{\tau}{2}}{[k\omega_s - 2\alpha(T_s - \tau)]\frac{\tau}{2}}\cos\left\{k\omega_s t - [\omega_0\tau - \tfrac{1}{2}k\omega_s(T_s - \tau)]\right\} \quad . \quad . \quad . \quad . \quad (19)$$

and a similar real harmonic component arises from the terms $F_3(\omega)$ and $F_4(\omega)$.

For our purpose, it will thus be sufficient to consider the functions for positive values of ω, since the negative values contribute nothing new.

(6.2) Amplitudes of Lines

Inspection of eqns. (13) and (15) shows that the spectrum has two maxima on the positive axis

at (i) $\omega = 2\alpha\tau = \omega_{B1}$ (say),
and at (ii) $\omega = 2\alpha(T_s - \tau) = \omega_{B2}$ (say).

Consider for the present the case when ω_{B1} and ω_{B2} are so far apart that the contributions of F_1 and F_3 can be considered independently. The maxima at ω_{B1} and ω_{B2} may or may not coincide with a particular line, depending on the value of τ relative to T_s. The zeros of the amplitude functions F_1 and F_3 occur when

$$\left.\begin{aligned}(\omega - 2\alpha\tau)(T_s - \tau)/2 &= \pm\,m\pi\\ \text{and}\ \ [\omega - 2\alpha(T_s - \tau)]\tau/2 &= \pm\,m\pi\end{aligned}\right\}\ \text{for } m = 1, 2, 3, \ldots \text{ but } not\ m = 0$$

$$\text{i.e. at} \qquad \omega = 2\alpha\tau \pm \frac{2m\pi}{T_s - \tau} \quad . \quad . \quad . \quad . \quad . \quad (20)$$

$$\text{and at} \qquad \omega = 2\alpha(T_s - \tau) \pm \frac{2m\pi}{\tau} \quad . \quad . \quad . \quad . \quad (21)$$

Now, let ω_{B1} or ω_{B2} be an integral multiple of ω_s. One line is now situated at the peak of a central maximum and successive lines are spaced at intervals ω_s $(=2\pi/T_s)$ about it. The zeros on the other hand are spaced at intervals $2\pi/(T_s - \tau)$ or $2\pi/\tau$ about the central maxima, respectively. Thus, only in the cases when $\tau = 0$ or $\tau = T_s$ do the lines and zeros coincide, and these cases are of little interest in a radar. Gnanalingam used a scheme for adjusting T_s at each range step, claiming that by reducing the beat note to the nearest whole number of cycles a single line spectrum was produced. It is easily seen that this result is due to the neglect of the period during each cycle when one of the tones (upper or lower) is not received. For short ranges (i.e. $\tau \to 0$) some value may accrue from using the method of synchronization proposed by Gnanalingam, but otherwise for values of τ of the same order of magnitude as T_s no advantage is obtained.

A scheme is proposed later to obtain some of the advantages of a coherent system, although it is not possible to reduce the beat-note spectrum to a single line.

The relative amplitudes of the two peaks are obtained by putting $\omega = \omega_{B1}$ and $\omega = \omega_{B2}$ in the appropriate functions F_1 and F_2, and one has

$$\frac{\text{Amplitude of lower beat-note}}{\text{Amplitude of upper beat-note}} = \frac{T_s - \tau}{\tau} \quad . \quad (22)$$

Fig. 2 illustrates by sketches the way in which the distribution changes with τ.

Fig. 2(*a*) shows the case where τ is small compared with T_s; the lower beat note then has a central maximum which is only slightly wider than two line intervals. The amplitudes of each peak in successive maxima on either side of the central one are, to a first approximation, in the ratio

$$\frac{2}{3\pi} : \frac{2}{5\pi} : \frac{2}{7\pi} \cdots \frac{2}{(2n+1)\pi}$$

to the central maximum which is a result independent of τ.

Thus the first peak on either side of the maximum is only 21% of the central peak.

Fig. 2(*b*) shows the case when $\tau = \frac{1}{2}T_s$. Although there is an

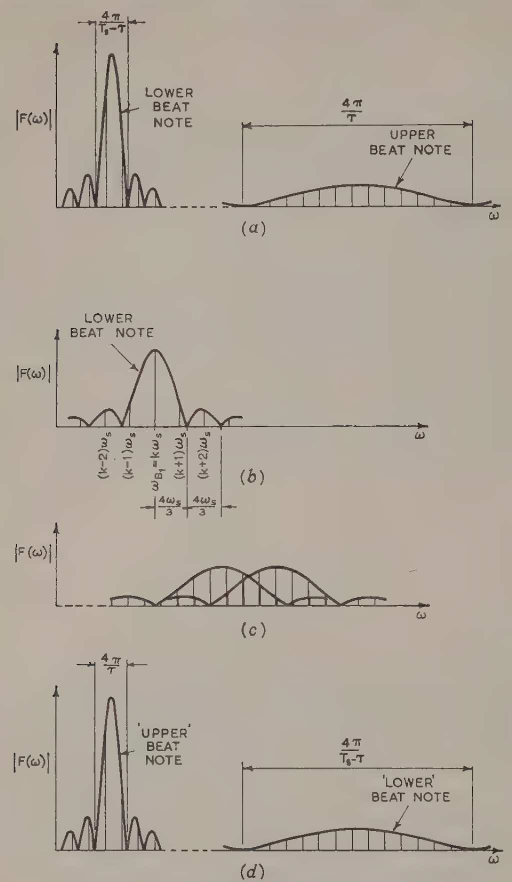


Fig. 2.—Distribution for selected values of τ.

(a) τ small.

Width of lower maximum $= \frac{4\pi}{T_s - \tau} > \frac{4\pi}{T_s}$

Therefore there are two or three lines between first zeros.

(b) $\tau = T_s/4$. Zeros spaced at $\frac{2\pi}{\frac{3}{4}T_s} = \frac{4\omega_s}{3}$

Therefore there are an integral number of cycles. But there are three lines in the first maximum and one line at the peak of the second (upper maximum not shown).

(c) $\tau \to T_s/2$.
Maxima nearly equal in amplitude. Lines overlapping.

(d) $\tau > T_s/2$.

integral number of harmonics in ω_{B1}, the spacing between first zeros is $8\omega_s/3$ and there are three lines in the first maximum. Furthermore the $(k \pm 2)$th lines on either side now occur at about the peak of the second maximum, and have amplitudes of about 21% of the kth line as shown above.

Fig. 2(c) shows the case when $\tau \to \frac{1}{2}T_s$. Quite clearly, the envelopes of the two sets of lines begin to merge, and to find the amplitude of the resulting oscillations we have to consider the phases of the two contributions, due to the upper and lower beat notes, respectively. Consideration of the argument leading to expression (19), and reference to the exact forms for F_1 and F_3, shows that there will be present in the receiver (video-frequency section) two oscillations both of frequency $k\omega_s$, one from the lower beat-note envelope with phase $(\omega_0\tau - \frac{1}{2}k\omega_s\tau)$ and the other from the upper beat-note envelope with phase $\frac{1}{2}k\omega_s(T_s - \tau) - \omega_0\tau$. Thus, if the two frequencies are superimposed and the resultant amplitude is measured, one has to take into account the difference of phase $2\omega_0\tau - \frac{1}{2}k\omega_s T_s$. Since $\omega_s = 2\pi/T_s$ the phase difference is seen to be just $(2\omega_0\tau - k\pi)$. Unless the transmitter possesses a high order of phase stability, the two components for each line will beat together and produce a fluctuating appearance which could be misinterpreted as a fading signal. This topic will be dealt with more fully in later Sections since it raises fundamental questions of the design of the system.

Fig. 2(d) shows the situation at still greater ranges. The upper and lower frequencies have changed places.

(7) EFFECT OF TARGET MOVEMENT

Section 2 described the system in relation to the ideal stationary reflecting target. However, if a target is receding and has a velocity component v in the line of sight, the relationships set forth above must be modified. If c is the velocity of propagation and R_0 is the range at $t = 0$, the delay τ is now a function of time:

$$c\tau = 2\left[R_0 + v\left(t - \frac{\tau}{2}\right)\right] \quad . \quad . \quad . \quad (23)$$

whence, as the velocity v is likely to be small compared with c,

$$\tau \simeq \frac{2}{c}(R_0 + vt) \quad . \quad . \quad . \quad . \quad . \quad (24)$$

The instantaneous beat-note frequency now consists of a pair of sliding tones, alternating in a progressively changing ratio, for the delay time τ_n and τ_{n+1} at the beginning of the nth and $(n+1)$th cycle of the transmitted frequency are related by the expression

$$\tau_{n+1} = \tau_n + \frac{2v}{c}T_s$$

The tones themselves may be obtained as before by direct differentiation of expressions (8) and (9), bearing in mind the fact that τ is no longer constant.

During time $\qquad -\frac{1}{2}T_s < t_n < -\frac{1}{2}T_s + \tau$

$$\frac{d}{dt}(\phi_g - \phi_e) = [\omega_0 + 2\alpha(T_s - \tau) + 2\alpha t_n]\frac{2v}{c} - 2\alpha(T_s - \tau) \quad . \quad . \quad . \quad . \quad (25)$$

and during time $-\frac{1}{2}T_s + \tau < t_n < \frac{1}{2}T_s$

$$\frac{d}{dt}(\phi_g - \phi_e) = (\omega_0 - 2\alpha\tau + 2\alpha t_n)\frac{2v}{c} + 2\alpha\tau \quad . \quad (26)$$

The corresponding positive values of the beat notes ω_{B1} and ω_{B2} thus both contain Doppler shifts, and these depend as expected on the instantaneous angular transmitted frequency; in fact, ω_{B1} and ω_{B2} are, respectively,

$$\left\{2\alpha(T_s - \tau) - \frac{2v}{c}[\omega_i + 2\alpha(T_s - \tau)]\right\} \text{ and}$$

$\left[2\alpha\tau + \frac{2v}{c}(\omega_i 2\alpha\tau)\right]$ (see Fig. 3). It is of interest that the Doppler component appears as a decrease in the beat-note frequency during the earlier, and as an increase during the later, part of the cycle.

The difference-frequency component, $\frac{1}{2}GV_eV_g \cos(\phi_g - \phi_e)$ in Section 5, will also, in general, have a non-repetitive waveform, and its Fourier transform F(ω) must therefore properly be interpreted as a continuous spectrum. The detailed analysis is beyond the scope of the present work. In most practical applications, however, it is adequate to think in terms of a slowly-changing line spectrum of the type described in Section 6. Even

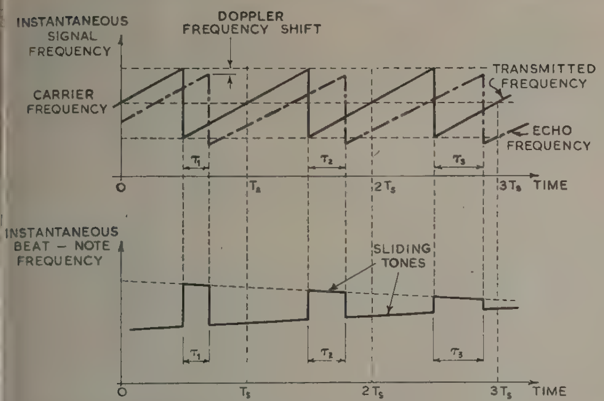


Fig. 3.—Production of the beat note when the echo comes from a moving target.

For a target receding at Mach 1, $\tau_{n+1} - \tau_n$ is of the order of only $2 \times 10^{-6}T_s$, so that τ will increase from 0 to T_s (thereby completing one cycle of spectral repetition) in a time of approximately $5 \times 10^5 T_s$; and at lower line-of-sight velocities the periodic time for τ will be proportionally greater.

(8) SCHEMATIC FOR A COHERENT SYSTEM

As shown above, the scheme proposed by Gnanalingam will not produce the desired result of a single-line beat note except in the trivial case of zero range. It is, however, possible to produce a coherent system in which each line of the beat-note spectrum is examined separately with an arbitrarily long time of integration equivalent to an ideally narrow pass-band filter. Fig. 4 shows a block diagram of an apparatus for doing this.

The swept frequency is generated by a sawtooth voltage which is itself triggered by a train of clock pulses at intervals T_s, thus ensuring that the duration of one sweep is accurately defined.

The same train of pulses is used to amplitude-modulate an externally generated intermediate frequency (for convenience this has been taken as 100 kc/s), thus producing side frequencies spaced at intervals of $1/T_s$ cycles per second. Each side frequency can be selected in turn by a variable-frequency high-Q-factor tuned amplifier, so that one has available a set of frequencies separated by the same intervals as the beat-note spectrum.

The receiver is shown in Fig. 4 only in outline. All linear amplifiers have been omitted. The beat-note frequencies and the output from the variable-frequency selective amplifier are passed into a balanced modulator, which produces the difference frequencies. Only one of these will be at exactly 100 kc/s, and this can be selected by the use of a phase-sensitive detector whose switching frequency is the unmodulated intermediate frequency. With an integration time as long as desired, one has the equivalent of a narrow-band filter, tuned to exactly 100 kc/s as suggested by Gnanalingam.[1] It will be noted that great stability of the intermediate frequency is not required, since it enters the system only as a 'carrier' of information, and is subtracted out again at the phase-sensitive detector. It is suggested that the intermediate frequency should be 100 kc/s, since that is a standard value for this type of work and techniques are therefore well known. Moreover, for the projected system, it falls well outside the possible range of beat-note frequencies so that second-channel interference is minimized and does not lead to confusion between, say, echoes from targets at ranges $3R$, $5R$ and higher-order modulation products arising from the beat note of a target at range R. This was a difficulty on which Gnanalingam remarked in Section 5.3 of his paper.

(9) RANGE AMBIGUITY

An f.m. c.w. radar ranging system of the type discussed suffers from a difficulty not encountered with pulse-type radars. Reference to Fig. 5 shows that for every range R_A there is a complementary range R_B, such that the upper and lower beat notes are identical. Thus the echo from range A has beat frequencies $4R_A\alpha/c$ and $2\alpha T_s - 4R_A\alpha/c$, and echoes from range B have beat frequencies $4R_B\alpha/c$ and $2\alpha T_s - 4R_B\alpha/c$. Ambiguity in range measurement will occur when

$$R_A + R_B = \tfrac{1}{2}cT_s \quad . \quad . \quad . \quad . \quad . \quad (27)$$

The obvious way of resolving this problem is to have available different sweep rates 2α, a method which is equivalent to a choice of values of the pulse-repetition frequency in a pulse radar. Another method of attack would be to use what may be called 'channel switching' to distinguish it from the first method, which is 'range switching'.

Fig. 5 has been drawn to show the special case when $R_A = \frac{1}{2}R_B$. In the general case, the two tones overlap for a fraction $\left(1 - \frac{4R_A}{cT_s}\right)$ of each cycle. Thus, there is in theory the possibility of confusion during virtually the whole cycle, but for

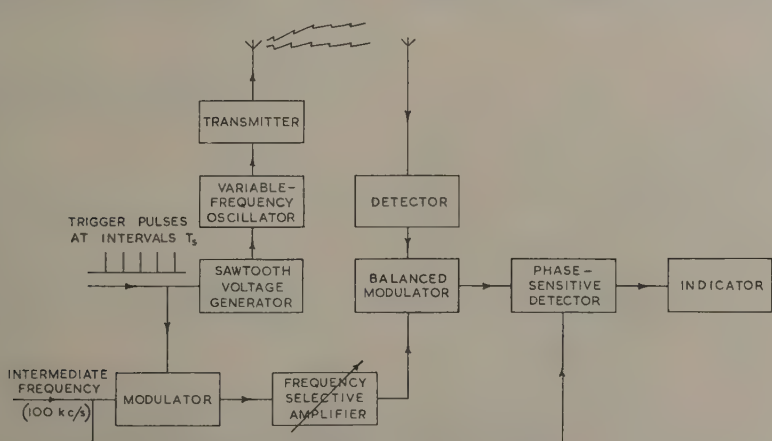


Fig. 4.—Schematic of a coherent system.

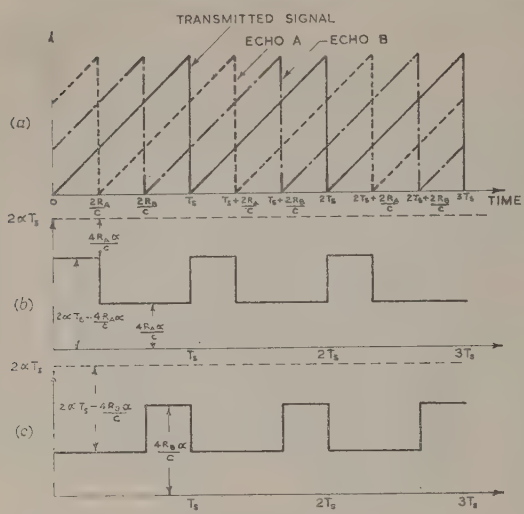


Fig. 5.—Production of complementary beat notes by ambiguous echoes.

$2R_A/c = T_s/3$

(a) Instantaneous frequency ω_t.
(b) Instantaneous beat-note frequency (echo A).
(c) Instantaneous beat-note frequency (echo B).

values of R_A small and tending to zero one must take into accoun the difference in amplitudes between the two signals. Even i only the normal radar range equation is applicable this confusion is not likely to arise for small values of R_A, since the corre sponding value of R_B is at extreme range. The ratio of the tw powers received will be

$$\frac{P_B}{P_A} = \left(\frac{R_A}{R_B}\right)^4 = \left(\frac{\tau}{T_s - \tau}\right)^4 \quad . \quad . \quad . \quad (28$$

Hence, if, for example, $R_A = \frac{1}{3}R_B$, the power ratio is $\frac{1}{81}$ so that the ambiguous signal is approximately 19 dB down on the signal from range A. For smaller values of R_A this ratio will be even greater. In a long-range equipment such as an ionosphere sounder, atmospheric attenuation will also increase signal difference by a considerable factor. For two comple mentary signals at nearly half range ($R_A = \frac{1}{4}cT_s$), however, th amplitudes will be more nearly comparable. As $\tau \to \frac{1}{2}T_s$ the time of overlap between the two tones, which is $\left(1 - \frac{4R_A}{cT_s}\right.$ of each cycle, tends to zero, i.e. the ambiguous tones from eac range occur at different times during the cycle; and this fac makes it possible to employ 'channel switching' to discriminat against the unwanted range in a way which is analogous to a pulse range gate. The analogy is not perfect, however, sinc some power from the complementary signal will get through

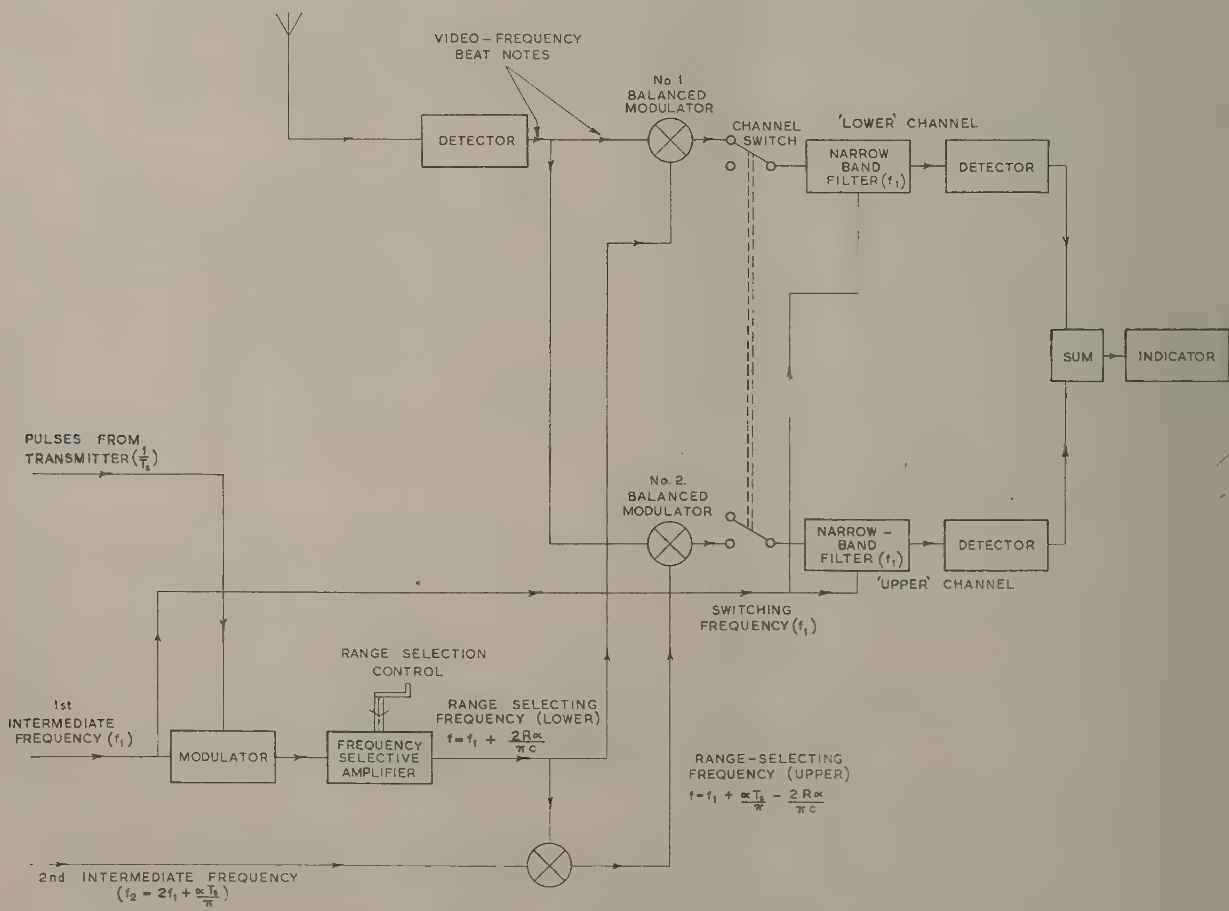


Fig. 6.—Schematic for a complete receiver with 'channel switching'.

but the improvement in ratio of desired to undesired signal strength will be greatest at the range where it is most needed.

Fig. 6 is a schematic of a receiver combining channel switching with a two-tone system for making use of the whole beat-note spectrum for one range. Once again, all linear amplifiers have been omitted for clarity.

In this scheme two 'intermediate frequencies' are required; the one at f_1 can well be 100 kc/s as before. The other, f_2, at $2f_1 + \alpha T_s/\pi$, will be required to beat with the upper maximum in the spectrum. In order to preserve coherence, both upper and lower range-selection frequencies are generated by the same modulation and selection process, using the principle of frequency-changing to provide separation in the frequencies. The two selecting frequencies are then made to heterodyne with the video-frequency beat note in the two balanced modulators. For a given range R two 100 kc/s tones will thus be produced, the one by difference between selecting frequency f_1 and the lower beat note, and the other by difference between selecting frequency f_2 and the upper beat note. The two frequencies are gated by a channel switch (with waveforms as shown in Fig. 7) and are then

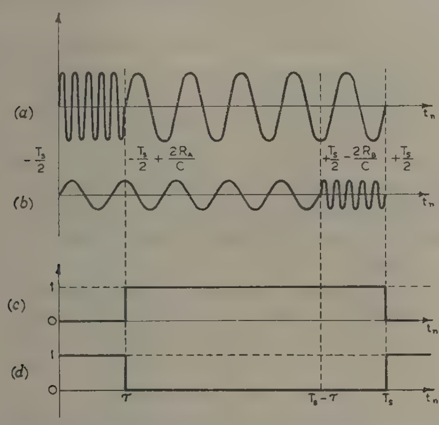


Fig. 7.—Gating waveforms for channel switching.
(a) Beat-note voltage; echo from range R_A.
(b) Beat-note voltage; echo from complementary range R_B.
(c) Lower-frequency gating waveform.
(d) Upper-frequency gating waveform.

passed to two separate narrow-band filters followed by phase-sensitive detectors. The channel switch is controlled by the setting of the variable-frequency selective amplifier. Thus, if the range selection control is set to receive signals from a range R_A (lower beat-note $\omega_{B1} = 4\alpha R_A/c$), the switch will change over during each sweep cycle at time $t_n = -\frac{1}{2}T_s$ and back at time $t_n = -\frac{1}{2}T_s + 2R_A/c$. Reference to Fig. 7 shows that the unwanted signal power from the ambiguous range will be multiplied by the factor $\left(1 - \frac{4R_A}{cT_s}\right)$.
Thus in the worst case when $R_A \simeq \frac{1}{4}cT_s$ and the two interfering signals are of comparable magnitude, the unwanted one is almost completely eliminated.

The realization of the channel switch should not prove difficult, since small errors in the switching instant will cause only a small deviation from the ideal case, and the degradation of the signal will not be greatly increased.

(10) RANGE ACCURACY AND DISCRIMINATION

(10.1) Range Accuracy

In a radar system of the type under consideration, the problem of measuring range is essentially that of measuring the time delay τ, for $\tau = 2R/c$. For a stationary target, the two tones ω_{B1} and ω_{B2} which comprise the beat note between echo and ground wave have angular frequencies $2\alpha\tau$ and $2\alpha(T_s - \tau)$; for a moving target these frequencies are, respectively,

$$2\alpha\tau + \frac{2v}{c}(\omega_i - 2\alpha\tau) \text{ and } 2\alpha(T_s - \tau) - \frac{2v}{c}[\omega_i + 2\alpha(T_s - \tau)]$$

Taking the first tone in each case,

$$R = \frac{c}{4\alpha}\omega_{B1}$$

for a stationary target

and
$$R = \frac{c}{4\alpha}\left(\frac{\omega_{B1} - 2v\omega_i/c}{1 - 2v/c}\right)$$

or, to a first order of approximation

$$R = \frac{c}{4\alpha}\omega_{B1} - \frac{v}{2\alpha}(\omega_i - \omega_{B1})$$

for one receding with line-of-sight velocity v. If ω_{B1} could be measured precisely, the effect of target movement would still appear as an apparent range decrease $\frac{v}{2\alpha}(\omega_i - \omega_{B1})$.

The remainder of the discussion will concern stationary targets. An ideally linear frequency sweep has been assumed throughout the paper and the analysis has been developed on that hypothesis. As 2α is the ratio of the total frequency excursion to the sweep duration, any inaccuracy in either of these parameters will give rise to a percentage error in α and therefore in range. Any departure from linearity would introduce perturbations in the modulation which would modify the spectral envelope.

A contribution to range error also arises from the fact that, whereas a change in τ causes the spectral envelope to move continuously in frequency, the lines are located at frequencies which are integral multiples of $2\pi/T_s$. Thus ω_{B1}, the peak of the envelope, will, in general, occur between two lines, and the problem to be solved is one of interpolation.

By increasing the bandwidth of the filter so that it passes two adjacent lines instead of a single line, the accuracy of interpolation may be improved by scanning the lines first singly and then in pairs. Nevertheless, a fixed-magnitude 'reading error' of the order of $\pm\pi/T_s$ will always be present in ω_{B1}, and therefore in the derived range, in addition to the percentage error arising from any uncertainty in α.

(10.2) Range Discrimination

In certain applications, a radar system deals primarily with extended targets, and range discrimination will be important only in so far as it is essential (as in an airborne altimeter) to assess accurately the range of the boundary of an extended target. Within the target itself, the echo range R and the corresponding delay time τ will be continuous, so that the spectrum will cover more or less uniformly a band of frequencies whose width depends on the difference between the maximum and minimum echo ranges for the target.

In other applications, the system may be called upon to discriminate between echoes from discrete objects at adjacent ranges, and it is the purpose of this Section to extend Gnanalingam's treatment of this topic.

He considered two targets whose echoes were of equal amplitude and assumed that the relative phases of their spectral lines were sufficiently random to justify the assumption that the average response V_s of two components V_{s1} and V_{s2} was given

by $V_s^2 = V_{s1}^2 + V_{s2}^2$. Eqn. (15) shows that, for stationary targets, at ranges having delays τ, $\tau + \delta\tau$, the corresponding phase of lines at frequency ω are $(\omega_0 - \frac{1}{2}\omega)\tau$ and $(\omega_0 - \frac{1}{2}\omega)(\tau + \delta\tau)$; thus there will be a phase difference $(\omega_0 - \frac{1}{2}\omega)\delta\tau$ between the interfering components. Beating will thus not be entirely random, but will be more or less severe according to the rapidity and extent of carrier fluctuations; pronounced beating will also occur, for a v.h.f. or h.f. carrier, for quite small differences $\delta\tau$ between the delays of the targets. Even in the idealized case when noise is absent, the problem is not readily susceptible of a mathematical analysis which is at the same time simple and rewarding; instead, typical cases are presented in Fig. 8, which shows alternative line spectra, with the upper and lower limits of beating, for target pairs of equal and unequal amplitude. In the left-hand spectrum of each pair, the range of the nearer target is such that a line occurs exactly at the peak of the spectral envelope; the right-hand spectrum of each pair shows an intermediate case.

Figs. 8(*a*)–(*d*) show, respectively, separations in ω_{B1} of

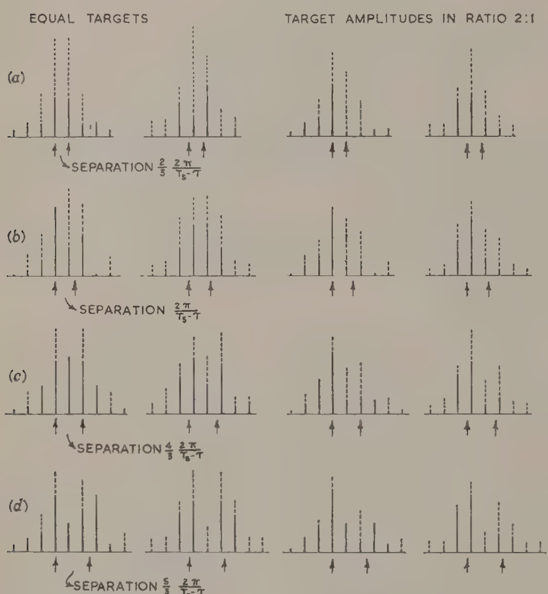


Fig. 8.—Typical line spectra for adjacent targets, showing upper and lower limits of beating.

The arrowheads below each line spectrum show the precise positions of ω_{B1} for the targets.

$\frac{2}{3}2\pi/(T_s - \tau)$, $2\pi/(T_s - \tau)$, $\frac{4}{3}2\pi/(T_s - \tau)$, and $\frac{5}{3}2\pi/(T_s - \tau)$; Fig. 8(*b*) corresponds to Gnanalingam's 'critical separation'. It is legitimate to conclude from the Figures that there is little probability of resolution for separation in ω_{B1} up to and including $2\pi/(T_s - \tau)$; the presence of beating indicates multiple targets but gives little help towards their resolution. The separation must exceed $2\pi/(T_s - \tau)$ by a significant factor before the positions of the individual targets can be confidently identified, and even then the line structure of the spectra may give only a relatively inaccurate estimate of the actual separation between the targets concerned [cf. Figs. 8(*c*) and (*d*)]. The probability of separation does not appear to be markedly altered if the amplitude of one target is reduced by half. In the presence of noise, integration times must be increased if the resolution is not to suffer.

The existence of the term $T_s - \tau$ in each denominator implies that range discrimination will deteriorate with increasing range; thus a separation in ω_{B1} of $2\pi/(T_s - \tau)$ is equivalent to a range separation $\delta R = \pi c/2\alpha T_s$ at $\tau = 0$, rising to $2\pi c/3\alpha T_s$ at $\tau = \frac{1}{4}T_s$ and to $\pi c/\alpha T_s$ at $\tau = \frac{1}{2}T_s$.

(11) CONCLUSIONS

The foregoing analysis of a frequency-modulated continuous-wave radar system shows that, except in the trivial cases of $\tau = 0$ and $\tau = T_s$, the beat-note spectrum cannot be reduced to a single line.

An alternative method has been proposed which does permit the detection, with adequately long time-constant, of individual lines in the beat-note spectrum. As a refinement, channel switching enables the additional ambiguity which occurs in the region of $\tau = \frac{1}{2}T_s$ to be minimized, and the use of a suitably chosen carrier frequency eliminates the possibility of confusion between echoes from targets at harmonically-related ranges.

Examination of the beating between nearby target echoes demonstrates the effect of carrier instability and suggests that, even under optimum conditions, the separation between the beat-note tones for two such targets must exceed $1/(T_s - \tau)$ cycles per second by a significant factor before the positions of the individual targets become clearly defined.

(12) ACKNOWLEDGMENT

The authors gratefully acknowledge the facilities made available to them at the Royal Military College of Science. They wish to thank the Dean for permission to publish the paper.

(13) REFERENCES

(1) GNANALINGAM, S.: 'An Apparatus for the Detection of Weak Ionospheric Echoes', *Proceedings I.E.E.*, Paper No. 1670, July, 1954 (**101**, Part III, p. 243).
(2) KEEP, D. N.: 'Frequency-Modulation Radar for Use in the Mercantile Marine', *ibid.*, Paper No. 1940 R, November, 1955 (**103 B**, p. 519).
(3) TUCKER, D. G.: 'Underwater Echo-Ranging', *Journal of the British Institution of Radio Engineers*, 1956, **16**, p. 243.
(4) KAY, L.: 'A Comparison between Pulse and Frequency Modulation Echo-Ranging Systems', *ibid.*, 1959, **19**, p. 105.
(5) WOODWARD, P. M.: 'Probability and Information Theory with applications to Radar' (Pergamon Press, 1953), p. 27.

621.391.821

The Institution of Electrical Engineers
Paper No. 3267 E
July 1960

EXTRA-TERRESTRIAL RADIO NOISE AS A SOURCE OF INTERFERENCE IN THE FREQUENCY RANGE 30–1000 Mc/s

By F. HORNER, M.Sc., Member.

(*The paper was first received* 16*th April, in revised form* 27*th August, and in final form* 10*th November,* 1959.)

SUMMARY

Published information on the intensity of noise from the galaxy, the sun and from other extra-terrestrial sources is presented in a form which shows their importance relative to the internal noise of typical receiving installations. Particular attention is paid to a half-wave horizontal dipole aerial and to an aerial with a 20° pencil beam directed horizontally.

(1) INTRODUCTION

When the various sources of interference to radio services over the whole frequency spectrum are studied, an important change is found to occur at a frequency in the region of 30 Mc/s. At lower frequencies atmospherics from thunderstorms are the main source of naturally occurring noise, and reception is normally limited either by atmospheric noise, by man-made noise or by interfering signals from transmitters. Above some frequency in the approximate range 10–30 Mc/s, depending on the epoch of the solar cycle, the season and the time of day, atmospheric noise is not propagated via the ionosphere from distant thunderstorm areas, and is important only when there are local storms. However, at these frequencies noise from sources outside the earth's atmosphere can penetrate the ionosphere and becomes the main source of natural interference for much of the time. The absence of regular long-distance propagation from transmitters reduces the interference from these, and reception is limited by extra-terrestrial noise, man-made noise or the inherent noise in the receiver.

From the work of radio-astronomers, knowledge of the radiation from extra-terrestrial sources has been greatly increased over the last few years. Several surveys of the subject are available,[1–3] but from the radio engineer's point of view the information is not in a convenient form for application to many reception problems. For example, one common requirement is a knowledge of the noise to be expected with a simple aerial system designed to receive transmissions from a fixed direction relative to the ground, but the distribution of noise sources is not usually described in terms of azimuth and elevation angle. Furthermore, the relative importance of extra-terrestrial noise and other types of interference at different frequencies requires clarification.

Most of the extra-terrestrial noise data at present available have been obtained with two general types of aerial—a $\frac{1}{2}\lambda$ horizontal dipole $\frac{1}{4}\lambda$ above the ground or a beam-type aerial with overall beam widths of from 1–2° up to about 20° in one or both planes. The first type integrates the noise from the whole of the visible hemisphere, except for directions nearly in line with the aerial, while with the second the spatial distribution of the sources is important.

The three types of noise which are discussed are the smooth background radiation (mainly from the galaxy), radiation from point sources and solar noise. The published data are discussed in relation to the problems of reception from a horizontal direction. It is assumed that the reception pattern of the directional aerial is unaffected by the presence of the earth, and the possible influence of side-lobes is largely ignored.

(2) UNITS AND CO-ORDINATE SYSTEMS

Noise intensity is often expressed as the effective aerial temperature T deg K. This may be converted to units in more common use in radio engineering by using the relationship that the noise power available from the aerial (into a matched load) is kTf_b watts, where k is Boltzmann's constant ($1 \cdot 37 \times 10^{-23}$ Joules/degree) and f_b (c/s) is the power bandwidth. The noise intensity is also expressed as $10 \log_{10} (T/300)$, which is the power available from the aerial, in decibels, relative to the thermal power at a reference temperature, arbitrarily chosen as 300° K. The effective aerial temperature is usually governed by the amplitude of one polarization component of the field. Since the noise is polarized nearly at random, the effective aerial temperature is practically independent of which component is measured.

Data on the distribution of sources other than the sun are nearly always expressed in galactic co-ordinates. The position of a source is defined on the celestial sphere, in relation to arbitrary reference points. For the present purpose it is necessary to replot the data to show the variations with time and direction relative to the earth's surface. Since the particular case of zero elevation angle is being studied, time and azimuth are the variable parameters. Now the sources of noise may be regarded as being fixed in space, so it is convenient to describe the temporal variations, not in solar time, but in sidereal time, which is based on the period of rotation of the earth relative to the celestial sphere. Sidereal time and local mean time (l.m.t.) change, relative to each other, by one day per year; noon l.m.t. corresponds to noon sidereal time at the autumn equinox, 1800 sidereal time at the winter solstice, and so on.

(3) THE SMOOTH BACKGROUND RADIATION

The smooth background radiation is received from the whole of the sky, with a pronounced maximum in the direction of the centre of the galaxy. With aerials of low or moderate directivity this background noise predominates over that from point sources and from the sun at most times.

A horizontal dipole integrates the noise over most of the visible hemisphere and the variations with time are not large. At 100 Mc/s, for example, the average aerial temperature is about 1 000° K and the noise power varies by less than ±2 dB with the movement of the galactic sources relative to the aerial pattern. Changes in the orientation and height of the aerial (provided that it is at least $\frac{1}{4}\lambda$ above the earth) change the pattern of the time variations, but do not greatly affect the average noise power or the extent of the variations. At fre-

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.
The paper is an official communication from the Radio Research Station, Department of Scientific and Industrial Research.

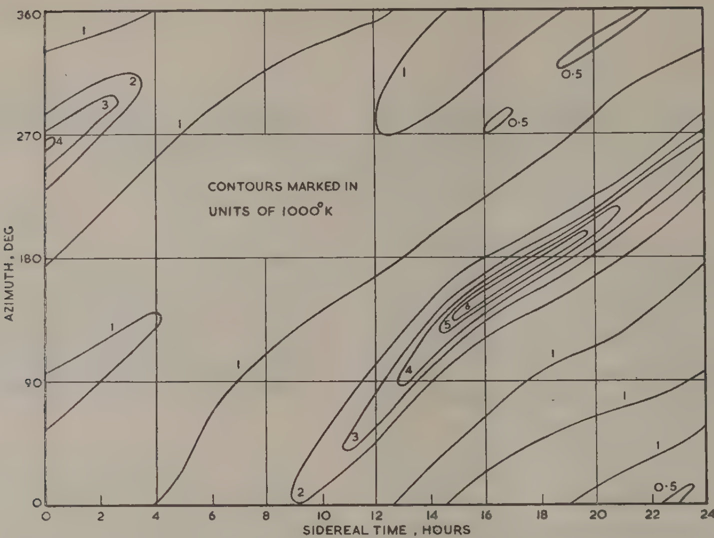


Fig. 1.—Effective temperature of an aerial due to the smooth background noise at 100 Mc/s (latitude 52° N).

quencies up to at least 200 Mc/s the aerial temperature is proportional to $f^{-2\cdot5}$.

Most data for directional aerials relate to beam widths of the order of 20° in both directions or to fan beams with this order of beam width in the vertical plane but narrower in the horizontal plane. As an example of how the data can be plotted in a convenient form, Fig. 1 shows the variation of the temperature, at 100 Mc/s, of a horizontally directed aerial at latitude 52° north as a function of azimuth (measured from north in the direction through east) and sidereal time. The data were derived from 81·5 Mc/s measurements, made at the Cavendish Laboratory, Cambridge,[4] and were converted to 100 Mc/s on the assumption that the frequency index was −2·5. The aerial beam was 15° wide vertically and 2° wide horizontally, but since the variation of noise with direction is smooth, similar values would be obtained with beams up to 20° × 20° with a slight reduction in the maximum aerial temperature in the direction of the galactic centre. One cross-section of Fig. 1 gives the variation of aerial temperature with time, at a given azimuth, and the other the variation with azimuth at a given time.

The frequency dependence of the background noise is shown in Fig. 2, in which the ordinate is the noise power available from the aerial expressed in decibels above the thermal noise available when the aerial temperature is 300° K. Fig. 2(*a*) refers to a $\frac{1}{2}\lambda$ dipole, $\frac{1}{4}\lambda$ above the earth, and Fig. 2(*b*) to a 20° pencil beam. In Fig. 2(*b*) curves are given for both average conditions and for the maximum noise from the galactic centre. It should be noted from Fig. 1 that a horizontally directed beam at latitude 52° north can receive maximum noise only from azimuths near to 180°.

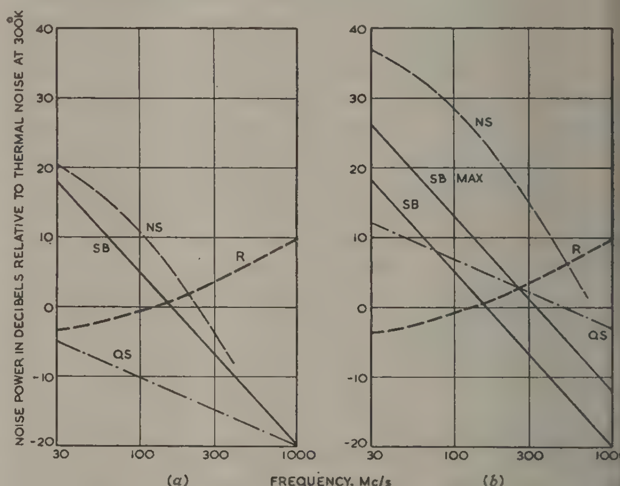


Fig. 2.—Comparison between various sources of noise.

R.—Receiver noise.
SB.—Smooth background noise (for the beam aerial the lower curve shows averag noise and the upper one maximum noise).
QS.—Noise from the quiet sun.
NS.—Approximate upper limit of noise during noise storms (noise outbursts ma be much more intense for short periods).
Note: The solar noise curves for the beam aerial are for an aerial pointing toward the sun.

(*a*) Horizontal $\frac{1}{2}\lambda$ dipole.
(*b*) 20° pencil beam.

(4) POINT SOURCES

The so-called point sources are mainly outside the galaxy and they all have very small angular width. With an aerial of low directivity, such as a $\frac{1}{2}\lambda$ dipole, the total radiation from these sources is negligible compared with the smooth background. With beamed aerials directed towards a point source the noise increases in importance as the beam is made narrower. However, if the beam width is not less than a few degrees, it never predominates over the background.

At the latitude of southern England (52° N), the most intense point source (in Cassiopeia) does not enter a horizontall directed 20° beam. The second most intense source, in Cygnu approaches the horizon in a northerly direction at certain time but even then the received power is only about one-tenth of tha from the galaxy in the same direction when a 20° beam is used Therefore, except with beam widths of only a few degree noise from point sources may be ignored as a source of inte ference. Such narrow beams are practicable with small aeria towards the upper limit of the frequency range under considera tion, but the noise is then small compared with the internal nois of currently available receivers (see Section 7).

(5) SOLAR NOISE

Solar noise consists of an almost constant component, partly ' thermal origin, with much larger bursts superimposed when ie sun is disturbed. Noise from the 'quiet' (undisturbed) sun n a $\frac{1}{2}\lambda$ dipole is small compared with the smooth background oise from the whole sky. With a 20° beam the maximum noise om the quiet sun is less than that from the galactic centre at equencies below about 300 Mc/s, but greater at higher fre-uencies. Solar noise differs from galactic noise, however, in iat the effective temperature increases rapidly as the beam idth is reduced, and if the beam width were equal to the ngular width of the sun, the effective aerial temperature would e 10^6 deg K at 100 Mc/s. Curves showing the maximum noise om the quiet sun are shown in Fig. 2.

Enhanced noise may be received when there is high sunspot ctivity, and these 'noise storms' may last for several days. The oise bursts may be up to 100 times more intense than the noise om the quiet sun at frequencies below 100 Mc/s, but the icrease is smaller at higher frequencies. Curves in Fig. 2 show ie maximum noise levels attained in a large proportion of orms, but much higher levels have been measured in exceptional orms. It can be seen that the levels with a $\frac{1}{2}\lambda$ dipole are some-hat in excess of the galactic noise levels, and that with a 20° eam directed towards the sun the maximum solar noise during storm may be much greater than galactic noise.

In addition to the noise storms, high noise levels are often ssociated with the occurrence of solar flares. These noise out-ursts may exceed the noise from the quiet sun by a factor of 0^4 or more. When they occur they are likely to be the pre-ominant source of noise at all frequencies, but they usually last or periods of only seconds or minutes.

Although maximum values of solar noise have been indicated, must be remembered that these can occur with a horizontally irected 20° beam only near sunrise and sunset, except at middle atitudes in winter. Unless a high degree of protection is equired, therefore, the important effects of solar noise are mited to times when the sun is disturbed, when noise bursts nay be received on side-lobes of the aerial.

(6) INFLUENCE OF THE IONOSPHERE

Over most of the frequency range considered, waves are trans-nitted through the ionosphere with little attenuation, and its nfluence may be neglected. It is only at frequencies just above he vertical-incidence critical frequency that ionospheric effects re important. Over a small range of frequencies, reception of xtra-terrestrial noise is possible only from a limited solid angle round the zenith, but this angle increases rapidly with frequency. ince the ionospheric effects depend in a complex manner on ime, frequency, the state of the ionosphere and the directivity f the aerial, they cannot be summarized conveniently. It may e noted that reception on a horizontally directed beam aerial s affected by the ionosphere up to higher frequencies than that vith any other type of aerial, but even in this case the main effects re confined to frequencies below 30 Mc/s.

7) RELATIVE IMPORTANCE OF EXTRA-TERRESTRIAL AND RECEIVER NOISE

The factors which must be taken into account in considering he influence of noise on a receiving system have been discussed y Norton and Omberg.[5] Present-day receivers of conventional lesign, based on such valves as disc-seal triodes, have noise igures only slightly below those quoted by Norton and Omberg. For comparison of receiver and external noise, curves have een added to Fig. 2 to show the contribution of the receiver itself to the total noise power. These curves are representative of the best currently available receivers of conventional design.

Important improvements in receiver sensitivity are being made, however, by the use of parametric and maser amplifiers.[7,8] These are still in the development stage, but have already resulted in receivers with effective noise temperatures well below those obtained with more conventional types. On the noise-power scale of Fig. 2, figures of 1 dB at 380 Mc/s[9] and 3 dB at 1 200 Mc/s[8] have been published for parametric amplifiers operating at room temperature, and the view has been expressed that it should be possible to obtain a value of -10 dB at 1 Gc/s with improvements in design.[8] A figure of -6 dB has been obtained in a maser amplifier at 450 Mc/s[10] by the use of refrigeration. When these low-noise amplifiers are in general use, extra-terrestrial noise will be an important design factor for most communication services, even at frequencies near 1 Gc/s.

In comparing the various sources of noise shown in Fig. 2 it should be borne in mind that the noise powers add directly. It is therefore necessary to convert from decibels to actual power as a step towards calculating the total power. However, any source of power which exceeds the others by at least 3 dB may be regarded as the predominant source.

It is convenient to summarize the data on the relative importance of the different sources of noise by dividing the frequency range into three parts, 30–100, 100–300 and 300–1 000 Mc/s, and discussing separately the half-wave dipole and the beam aerial. It is assumed that the receiving system is well designed.

(7.1) Half-Wave Dipole

In the lowest frequency range the background radiation is normally predominant, but solar noise may exceed this during noise storms. Noise bursts may exceed the background noise by a large factor for short periods. Noise from the quiet sun can be neglected.

The middle range is transitional; at 100 Mc/s external noise is predominant, but at 300 Mc/s receiver noise is the normal limitation with conventional receivers, except possibly during a solar noise outburst. New types of low-noise receiver may have noise levels slightly below the smooth background and noise during storms at 300 Mc/s, so that external noise will predominate over the whole of the middle range.

In the highest frequency range the receiver is always the limiting factor, if of conventional design, and even the new low-noise amplifiers are likely to have noise levels exceeding the aerial noise at 1 Gc/s. Noise from solar outbursts will, however, be predominant for short periods.

(7.2) 20° Beam Aerial

In the lowest range the background noise greatly exceeds receiver noise at all times and particularly when the beam is directed towards the galactic centre. The maximum noise from the quiet sun is somewhat below the background noise. During noise storms the maximum solar noise may be much more intense than even the noise from the galactic centre.

With conventional receivers the smooth background noise is generally below receiver noise over most of the middle frequency range, but above it if the galactic centre is in the beam. With new types of amplifier it may be possible to reduce receiver noise below the general level of the smooth background over all this range. Maximum noise from the quiet sun is comparable with that from the galactic centre over this range.

In the upper frequency range all extra-terrestrial noise is below the noise level of conventional receivers except the maximum noise from the disturbed sun. Considerable improvement in the quality of amplifiers will be required before the general level of the smooth background noise will become significant, but

parametric and maser amplifiers are likely to reduce receiver noise below the maximum noise from the quiet sun and possibly below that from the galactic centre.

Over the whole frequency range 30–1 000 Mc/s maximum solar noise during noise storms exceeds all other noise except that in conventional receivers above 600 Mc/s. Even when the sun is not in the main beam, this noise may be significant in a practical aerial system having side lobes. Large solar noise bursts will also be received with the sun off the main beam.

Although this survey is concerned with relatively simple aerials with which the beam width is unlikely to be less than 20°, a few properties of narrower beams may be noted. The effective temperature of the aerial due to solar noise when the sun is in the centre of the beam will be roughly inversely proportional to the cross-sectional area of the beam. Maximum noise from the galactic centre will also be increased, but not to the same extent. The effect of reducing the beam width is, therefore, to increase the importance of solar noise, but for shorter periods of time. When the sun is not in the beam, galactic noise is likely to predominate except at the highest frequencies. Some of the main point sources will be significant with beam widths of the order of 1°.

(8) CONCLUSIONS

The discussion is intended to show, in a general way, the conditions in which extra-terrestrial sources of noise must be taken into consideration in the design of a radio service. The actual design must take into account the directivity of the aerial, including the effects of side-lobes (with a beam aerial) and such operational considerations as the extent to which short interruptions can be tolerated or minimized by planning. These are factors which must be treated statistically in relation to the particular application.

The data discussed show that with aerials of only moderate directivity the use of new amplifying techniques is likely to make extra-terrestrial noise an important factor over the whole frequency range under discussion. The use of aerials more directive than those discussed will lead to more serious interference, but for shorter periods.

(9) ACKNOWLEDGMENT

The above review was carried out as part of the programm of the Radio Research Board and is published by permission o the Director of Radio Research of the Department of Scientifi and Industrial Research.

(10) REFERENCES

(1) Pawsey, J. L., and Bracewell, R. N.: 'Radio Astronomy (Clarendon Press, Oxford, 1955).

(2) Brown, R. H., and Lovell, A. C. B.: 'The Exploration o Space by Radio' (Chapman and Hall, London, 1957).

(3) Ko, H. C.: 'The Distribution of Cosmic Radio Backgroun Radiation', *Proceedings of the Institute of Radio Engineer* 1959, **46**, p. 208.

(4) Baldwin, J. E.: 'A Survey of the Integrated Radio Nois Emission at a Wavelength of 3·7 metres', *Monthl Notices of the Royal Astronomical Society*, 1955, **11** p. 684.

(5) Norton, K. A., and Omberg, A. C.: 'The Maximum Rang of a Radar Set', *Proceedings of the Institute of Radi Engineers*, 1947, **35**, p. 4.

(6) Tremlett, C. A., and Williams, A. D.: 'Triodes an Tetrodes for UHF–SHF Operation', *Electronic Enginee ing*, 1958, **30**, p. 335.

(7) Weber, S.: 'The Mavar: A Low-Noise Microwave Ampl fier', *Electronics* (Engineering Issue), 1958, **31**, p. 65.

(8) Heffner, H.: 'Solid-State Microwave Amplifiers', *Tran actions of the Institute of Radio Engineers*, 1959, **MTT–** p. 83.

(9) Engelbrecht, R. S.: 'A Low-Noise Nonlinear Travellin Wave Amplifier', *Proceedings of the Institute of Rad Engineers*, 1958, **46**, p. 1655.

(10) Kingston, R. H.: 'A UHF Solid-State Maser', *Transactio of the Institute of Radio Engineers*, 1959, **MTT–7**, p. 9

(11) Hausman, A. R.: 'Dependence of the Maximum Range Tropospheric Scatter Communications on Antenna ar Receiver Noise Temperatures', *ibid.*, 1958, **CS–6**, p. 35.

1.316.84

The Institution of Electrical Engineers
Paper No. 3277 E
July 1960

THE LONG-TERM STABILITY OF FIXED RESISTORS

By H. F. CHURCH.

(*The paper was first received* 21*st December*, 1959, *and in revised form* 7*th April*, 1960.)

SUMMARY

The causes of long-term failure under practical conditions of use or orage of different types of fixed resistors commonly used in electronic uipment have been investigated. Some reported life tests have proeded without interruption for almost four years. Carbon-composion (grade 2) resistors under load fail by slow thermal degradation of e resistive material. Drift of value may also occur if unloaded sistors of this type are stored in a damp atmosphere. Vitreousamelled wire-wound resistors made with fine wire may fail during opical exposure both unloaded and especially when lightly loaded th direct current. This is owing to electrochemical corrosion taking ace at faults in the vitreous coating. High-stability cracked-carbon rade 1) resistors may fail rapidly under light d.c. load by electroemical action if moisture condensation occurs and the protective int or varnish coating is inadequate.

Tests for long-term resistor stability are critically discussed.

(1) INTRODUCTION

Fixed resistors for use in electronic equipment are usually quired to maintain their resistance over a very long period of peration or storage. Ultimate failure may be sudden, as in e case of wire-wound resistors which usually open-circuit with tle warning, or there may be a slow drift of value which entually affects the proper functioning of the circuit in which e resistors are connected. The latter type of failure commonly curs with carbon-composition (grade 2) resistors: when opencuit occurs with these resistors it is usually subsequent to a adual lowering of value, over-heating because of the increased rrent and eventual thermal instability and burning away of e organic matter.

Grade 1 (high-stability) carbon resistors behave more like the ire-wound type in this respect: drift of value is relatively small ıt when failure occurs it is a rapid open-circuit.

The amount of drift which can be tolerated before resistors n be considered to have failed obviously depends upon their nction in the circuit.

The Electrical Research Association has, over a number of ars, made a detailed study of the causes of deterioration of veral commonly available types of resistors, e.g. mouldedrbon composition, both insulated and uninsulated, carbonsin film and vitreous enamelled wire-wound. Some aspects failure of high-stability pyrolytic-carbon-film resistors have so been investigated.

This paper describes the investigations made on the different pes and discusses the implications of the results obtained. It based on a series of E.R.A. Technical Reports.[1–7]

Failure, i.e. excessive deviation from the nominal value, may cur when the resistors are operated at full-voltage rating or ıring storage, or it may be especially rapid under conditions light loading. All these aspects of failure are considered.

The object of the work was to study the causes of long-term terioration of resistors and not to compare the performance different makes. Since some of the investigations started as long ago as 1952 and many of the tests have taken place over periods of several years, the resistors dealt with should not be regarded as necessarily representative of recent manufacturing practice. In fact, it is known that in several instances recent changes have been made in manufacturing methods, e.g. the use of improved vitreous coatings for wire-wound resistors and better protective varnishes for deposited-carbon-film resistors.

The paper is concerned only with long-term stability. Other characteristics of fixed resistors are comprehensively reviewed by Dummer,[8] who also deals with manufacturing methods and testing techniques, and includes an extensive bibliography.

(2) CARBON-COMPOSITION RESISTORS (GRADE 2)

(2.1) Moulded Resistors

The resistive element in this type is usually made by pressing into the form of a rod a powder consisting of a finely-divided carbon conductor, an inert inorganic filler and a synthetic-resin binder. After pressing at room temperature, the resistors are heat-treated to cure the resin and are finally wax-impregnated to fill the pores. Non-porous bodies may be formed by hotpressing; some experimental resistors of this type have also been studied.

Connections to moulded carbon-composition resistors may be moulded in, or the ends may be copper sprayed and a metal cap pressed on or wires soldered on. With insulated types the rod is enclosed in a ceramic or synthetic-resin sleeve.

Of the cold-pressed types, two makes, A and B, both normal commercial products, were investigated and, in general, behaviour was similar, differences being a matter of degree rather than kind. With both makes the resistance depends upon a series of imperfect contacts between carbon particles or agglomerates and is greatly influenced by volume changes or deformation caused by swelling in liquids or vapours, or by mechanical stress.

Exposure of uninsulated resistors to organic solvent vapours caused a marked increase in resistance, although gradual recovery to the original value took place on subsequent exposure to normal atmosphere. With A, in both acetone and benzene vapour an increase of 4 to 5 times was observed, in vapour from silicone fluid (viscosity 0·5 cS), 40–50%, and over hot transformer oil, 20%. With B in benzene and acetone vapour changes were observed and found to be somewhat smaller than with A, but in the same direction.

Swelling of the resin and an increase in resistance is also caused by moisture. Resistors of both makes were first dried to constant resistance and then conditioned for 5000–6000 h at 85% relative humidity. The increases observed in moist atmosphere were

	Increase, %
Make A. 1 watt 100 kilohms, wax-impregnated ..	4
Make A. 1 watt 100 kilohms, unimpregnated	10
Make B. 1 watt 100 kilohms, wax-impregnated ..	15
Make B. 1 watt 100 kilohms, unimpregnated	24
Make B. $\frac{1}{2}$ watt 100 kilohms, insulated	2

The unimpregnated resistors were specially supplied by the makers for the investigation and are not commercially available.

Written contributions on papers published without being read at meetings are ited for consideration with a view to publication.
The paper is based on Reports Refs. Z/T92, Z/T95, Z/T96, Z/T105, Z/T107, Z/T113, d Z/T124 of the British Electrical and Allied Industries Research Association.

The increases for these were more rapid as well as of greater magnitude than for the impregnated types. The insulated resistors were protected by a ceramic tube cemented at the ends. This evidently excluded or greatly reduced the rate of access of moisture to the resistive element. In all cases, the effects of moisture were reversible or nearly so.

Resistors which had deteriorated through ageing under load or had been over-stoved were more sensitive to moisture than normal ones.

The experimental hot-moulded resistors (make C) were also moisture-sensitive. A 5 000 h exposure of dried resistors to 85% relative humidity caused a mean increase in resistance of 16%. Resistance was still increasing slowly after 15 000 h.

Some special tests on small ($\frac{1}{4}$ watt) low-value resistors (less than 100 ohms) which had increased considerably in value, in some cases by more than 90% over a number of years whilst stored in sealed bags made from 5 mil polythene sheet, showed that there were two causes for the failure. First, there was an increase due to change in the resistance of the rod itself, as above, and secondly, in these particular resistors which are now of an obsolete design, a high contact resistance due to loosening of the embedded wire connection resulted from the swelling of the moulded body. Fig. 1 shows average change of value for

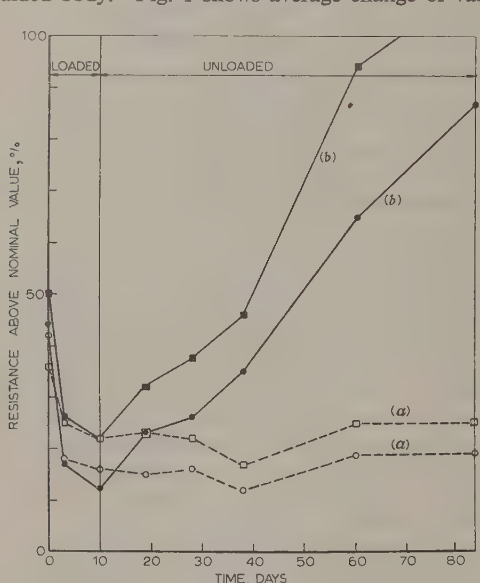


Fig. 1.—Changes in resistance of grade 2 resistors on loading and subsequently exposing unloaded to (*a*) laboratory atmosphere or (*b*) 100% relative humidity.

□, ■ 47 ohms. ○, ● 82 ohms.

resistors on drying out by loading for 10 days and subsequently storing in (*a*) normal room atmosphere and (*b*) in air at 100% relative humidity. Each plotted point is a mean determination on three typical resistors. Comparison of curves (*a*) and (*b*) clearly illustrates the effects of moisture. Changes in excess of 15% could be avoided by improved design of the contacts and manufacturers now claim to have done this. It was also evident from these tests that sealing in thin polythene bags does not exclude moisture indefinitely; appreciable permeation of moisture through thin plastic sheeting during long storage in moist atmosphere is expected on theoretical grounds.[3]

Longitudinal compression of resistor rods of make A was found to cause a linear and reversible decrease of resistance of 1% per 100 kg/cm² for loading up to about 165 kg/cm². Beyond this, permanent deformation occurred, resistance increased, ar ultimately the resistor shattered at 700 kg/cm². With B resisto the rate of decrease of resistance with load was the same but th changes remained linear and reversible up to 1 200 kg/cm², beyor which collapse of the material occurred suddenly.

(2.2) Carbon-Resin Film Resistors

Carbon-resin film resistors (make D) consist of a glass tul on the outside of which is formed a resistive film composed graphite or carbon black bonded with a mixture of synthet resins. The film is applied to the tube as a varnish which is the cured by heating. The terminal wires project into the glass tul and are cemented to it with a conductive cement which als serves to connect the wires to the resistive film. The element finally embedded in mica-filled phenolic resin which insulat the film and protects it from mechanical damage. Resisto are sometimes dipped in molten wax under atmospheric pressu to enable them to pass certain high-humidity tests.

Carbon-resin film resistors are sensitive to moisture in muc the same way as other grade 2 resistors. Unwaxed 47-kiloh $\frac{1}{2}$-watt resistors increased in value by 10 to 15% when stored f 1 000 h at room temperature in air almost saturated with wat and appeared to be still increasing in value. With waxe resistors the rate of increase in value was less than half that f unwaxed ones.

(2.3) Life under Load

Endurance tests have been conducted on all the vario grade 2 resistors described in Sections 2.1 and 2.2. Differe values of makes A, B and C were tested but with make D li tests were confined to 47-kilohm values. Tests by the latt manufacturer, however, showed no important differences performance between different values in the range 470 ohm 4·7 megohms.

Tests were carried out (*a*) with continuous nominal full loadi at an ambient temperature of 70° C and (*b*) at room temperatu with twice full loading for makes A, B and C and 1·8 times fu loading for make D. Tests (*a*) and (*b*) were expected to l approximately equivalent since the running temperature of t resistors was found to be roughly the same under the two co ditions. The internal temperature of the resistors under loa was measured by drilling a small radial hole and inserting a fin wire thermocouple.

In most cases comparisons between a.c. and d.c. loading we made, and with make D the effects of intermittent loading we also investigated.

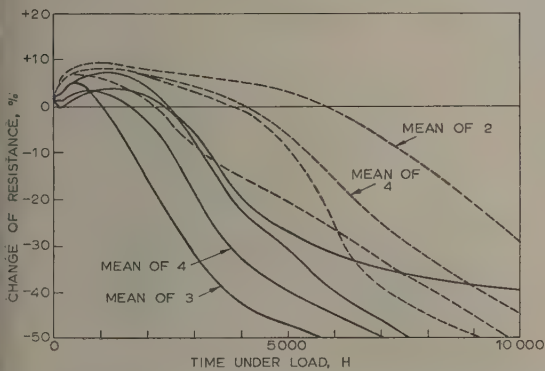
In all these tests the power dissipation in individual resisto was kept constant by manual adjustment of applied voltage. this way deterioration could be followed indefinitely witho thermal instability. During the tests resistances were measur from time to time under load. With some resistors weight a dimensional changes were also measured during the course ageing.

(2.3.1) Comparison of A.C. and D.C. Stress.

Comparative life-test results using d.c. and a.c. stress (50 c were obtained on make A at 70° C and 20° C ambient and make D at 20° C. No significant difference in performan between a.c. and d.c. loading, for the same power dissipatio was observed over periods of more than 10 000 h. It was co cluded, therefore, that later full-load tests on grade 2 resisto need only be carried out under a.c. stress. Electrochemic effects are evidently absent in grade 2 resistors under hot runni conditions.

This equivalence of alternating and direct current does n apply to high-stability carbon resistors, which may be mu

orter lived on direct current. In fact, in the endurance test ecified in B.S. 2112: Part 1: 1959 for fixed carbon resistors, a ect voltage is specified for high-stability resistors although her a direct or an alternating voltage is permitted for normal-bility resistors. There is a similar requirement in Inter-rvices Specification RCS112 for carbon-composition fixed sistors.



g. 2.—D.C. life tests on 10-kilohm grade 2 fixed carbon-composition resistors of make A.

---- Loaded to 1 watt at 70° C.
—— Loaded to 2 watts at room temperature.

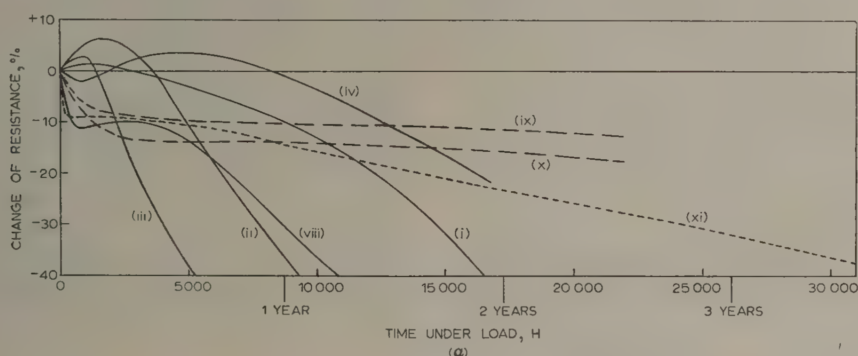
(2.3.2) **Comparison of Different Makes and Types.**

Resistance/time curves for nominally identical resistors showed considerable variability. Fig. 2 shows sets of life curves for 1-watt 10-kilohm resistors of make A loaded at both 1 watt (direct current) at 70° C and 2 watts at room temperature. For the sake of clarity of reproduction, closely similar curves have been combined as single mean curves.

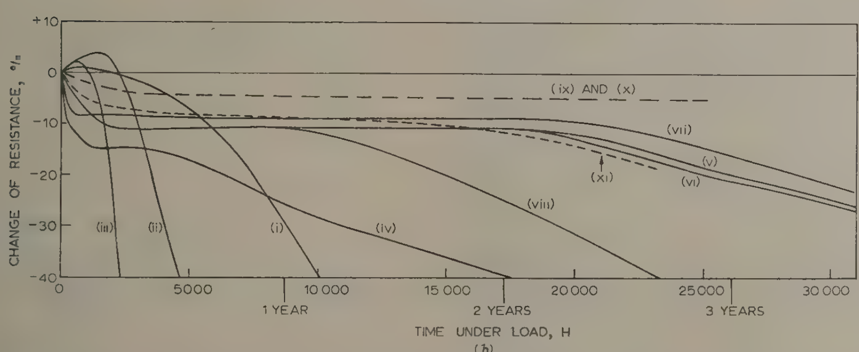
Standard deviation of resistor life, where life is defined as the time for the resistance to change by one-fifth, varies with different makes, values and loadings. The coefficient of variation ranges from 15 to 60% and averages 30% at both 20° C and 70° C.

Mean life test curves for commercial resistors of different makes run continuously at nominal rating at 70° C, and at twice this rating at room temperature in air (1·8 times the 70° C rating for make D), are shown in Figs. 3(*a*) and (*b*). These show mean changes in groups of nominally identical resistors, the group size varying from six to twelve according to the make and resistance.

Behaviour during the first 1 000 h or so is different with different makes. Make A all rise by a few per cent during this period, and then fall steadily, ultimately attaining very low values. Make B in general fall by 10–15% during the first 500–1 000 h, remain steady for some thousands of hours and then fall again to reach very low values eventually. Make C fall similarly during the first 1 000 h or so and then remain almost steady for at least 20 000 h. Make D fall rapidly by 10%, then less rapidly for some thousands of hours, but eventually (after



(*a*)



(*b*)

Fig. 3.—Long-term endurance of grade 2 fixed carbon-composition resistors under continuous a.c. load.

(*a*) At 70° C at rated wattage.
(*b*) At 20° C at twice 70° C rating for makes A, B and C, and at 1·8 times 70° C rating for make D.

(i) Make A, uninsulated, 1 watt, 100 Ω.
(ii) Make A, uninsulated, 1 watt, 10 kΩ.
(iii) Make A, uninsulated, 1 watt, 100 kΩ.
(iv) Make B, uninsulated, 1 watt, 100 kΩ.
(v) Make B, insulated, ½ watt, 100 Ω.
(vi) Make B, insulated, ½ watt, 1 kΩ.
(vii) Make B, insulated, ½ watt, 10 kΩ.
(viii) Make B, insulated, ½ watt, 100 kΩ.
(ix) Make C, insulated, ¼ watt, 100 kΩ
(x) Make C, insulated, ¼ watt, 220 kΩ.
(xi) Make D, insulated, ½ watt, 47 kΩ.

10 000 h) again fall more quickly. The early fall shown by most resistors is usually due to drying out, and is mostly recoverable on return to equilibrium with room conditions without load. It will be seen from Fig. 3(*a*) that even the best resistors of this type have an effective life of only a few years under rated conditions, i.e. with full load applied continuously at 70° C ambient. In practice, however, they are rarely maintained indefinitely under full load and at maximum category temperature, so considerably longer lives are generally obtainable.

Changes of weight and dimensions of some make A resistors (uninsulated solid rods) were measured throughout life under load. 1-watt resistors (70° C rating) loaded to 2 watts at room temperature lost weight continuously during life. After 4 000 h, average loss of weight was 3 % and was then continuing at about 0·5% per 1 000 h. Weight changes were of the same magnitude whether the resistors were wax-impregnated or not.

Shrinkage in length and diameter also occurred, the linear change being about 0·5% after 2 000 h. This shrinkage draws the carbon particles into closer contact and accounts for the eventual reduction in resistance.

(2.3.3) Intermittent Loading.

In addition to continuous a.c. and d.c. load tests, resistors of make D were subjected to intermittent a.c. loading with on- and off-load periods both of 2 min duration. The loading was at room temperature and to the same wattage as in Section 2.3, condition (*b*).

Intermittent loading was somewhat more harmful than continuous loading, the comparisons being made after the same total period of loading (up to 10 000 h), but the differences were hardly significant.

(2.3.4) Effects of Change in Manufacturing Technique.

Life may be considerably influenced by variations in manufacturing technique. For example, behaviour under load is very dependent on the type of carbon dispersed in the resin and on the state of cure of the resin.[9]

With moulded carbon-composition resistors it has been found that wax impregnation of the cured body has a considerable effect on endurance. With make A, life under load was very short when wax impregnation was omitted, and for longest life impregnation under vacuum is necessary. Also, the nature of the wax plays a significant part. These effects are illustrated by the mean life curves of Fig. 4. Wax impregnation is also important with make B resistors; its omission may reduce their life to one-fifth. The observed short lives with unimpregnated resistors were the same whether the resistors were specially supplied unimpregnated by the manufacturer or the wax impregnant was extracted from impregnated resistors by a solvent before carrying out the life tests.

It is evident that the wax plays an important part in delaying contraction of the resin by heat. Oxidation of the wax appears to diminish its effectiveness: when life tests on impregnated resistors are carried out in an atmosphere of nitrogen instead of air, or when air is excluded by provision of an insulating sleeve, life is much extended. Resistors of make A were six times longer lived in nitrogen than in air, whilst uninsulated resistors of make B were four times longer lived in nitrogen. Insulated resistors of make B were some five times longer lived under full-load conditions than the uninsulated type, the difference presumably being due to the protection afforded by the ceramic covering.

With make D, although the carbon-loaded resin-film resistive element is not wax-impregnated, the moulded plastic covering plays a considerable part in controlling stability. First, it reduces the running temperature of the element under load by

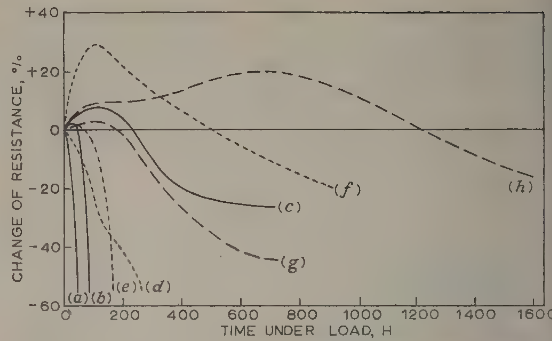


Fig. 4.—Endurance of 100-kilohm grade 2 resistors of make A variously impregnated.

(*a*) Unimpregnated.
(*b*) Imperfectly impregnated with manufacturer's wax, m.p. 82° C.
(*c*) Vacuum-impregnated with manufacturer's wax, m.p. 82° C.
(*d*) Vacuum-impregnated with paraffin wax, m.p. 57° C.
(*e*) Vacuum-impregnated with cold-curing polyester resin.
(*f*) Vacuum-impregnated with synthetic long-chain-ester wax, m.p. 141° C.
(*g*) Vacuum-impregnated with wax 'B', m.p. 100° C.
(*h*) Vacuum-impregnated with wax 'X', m.p. 79° C.

All groups were loaded at 2½ watts at room temperature.
All curves represent mean behaviour of six resistors.

increasing the rate of heat dissipation; the complete resisto[r] could be operated at nearly twice the wattage of uncovere[d] elements for the same running temperature. Secondly, eve[n] with the wattage adjusted so that covered and uncovere[d] resistors ran at the same temperature, life was some five tim[es] longer with the plastic covering.

(2.4) Effects of High Temperature without Electrical Loadi[ng]

The effects of storage of makes A and B in air at 105° C— the running temperature at full loading—are shown in Fig. [5]. The effects are similar to those of electrical loading with bot[h] makes, although rather less severe. Initial changes during th[e] first few hundred hours and subsequent major deterioration occ[ur] as with electrical loading, and the same wide spread of resul[ts] was observed.

Similarly, resistors of make D maintained continuously a[t] 120° C, the temperature of the resistive element when fully loade[d] behaved as fully-loaded resistors in air at rated ambient temperature (Fig. 5).

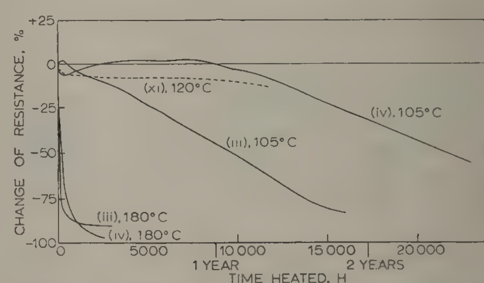


Fig. 5.—Effect of storage of grade 2 resistors at elevated temperatu[re] without load.

The numbers on the graphs have the same meaning as in Fig. 3.

Makes A and B when heated continuously at still higher temperature (180° C) deteriorated very rapidly, in both cases th[e] resistance falling to a small fraction of the initial value in a fe[w] hundred hours (Fig. 5). It is evident from this that resistanc[e] is very dependent on the period of stoving or curing, which [is] carried out at a temperature of about 230° C.

It is apparent from these experiments that ageing under loa[d]

resistors of the carbon-composition type (grade 2) is essentially thermal process, i.e. gradual thermal degradation of the ;sistive element occurs whether the heating is by electrical ading or is externally applied, and this leads to progressive duction in resistance. Initial changes, i.e. those taking place uring the first 500–1 000 h, give no indication whatever of ultiate life. For example, the shortest-lived group of resistors in g. 3 showed almost no deviation from the initial value after)00 h, whereas some of the longest-lived samples showed a)% drop in resistance after 1 000 h.

(2.5) Overload Tests

Endurance tests under conditions of overload were carried ut on several makes of grade 2 resistors and were of interest om two points of view:

(*a*) It may be possible to accelerate life so that prediction of the fects of changes in manufacturing methods are obtainable in a latively short time.

(*b*) In applications where long life is not required, but stability ver a relatively short period of operation is necessary, the iformation is of direct interest.

Fig. 6 illustrates the effects of overload for various grade 2 :sistors. These overload tests were performed at room temerature, normal full load (0% overload) under these conditions eing taken as twice the 70° C rating for makes A and B and ·8 times this rating for make D (see Section 2.3). It is evident hat the makes tested are not equally affected: 50% overload aused a mean decrease of life to roughly one-fortieth for makes A nd C, but to only one-seventh for make B. There is thus no eneral method of estimating life from the results of overload :sts. The correction factor must be determined experimentally or each type or make. In all these cases 'life' corresponds to 20% deviation from the initial resistance under load. Life ssessed in this way is uninfluenced by instantaneous, usually eversible, changes due to temperature.

(3) VITREOUS-ENAMELLED WIRE-WOUND RESISTORS

The wire-wound resistors investigated consisted of a singleayer winding of superfine nickel-chromium wire (diameter 0·000 6–0·001 2 in) on a ceramic tube, connections being made by welding on to a metal end cap or brazing on to a stout wire. The winding was protected by a smooth coating of vitreous enamel. Resistors made with these fine wires mostly fell within the range 10–100 kilohms.

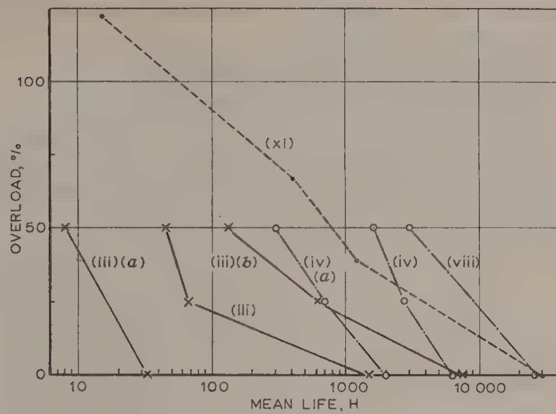


Fig. 6.—Overload tests on grade 2 fixed carbon-composition resistors.

'Life' is defined as the mean time for the resistance to change by 20% from the initial value under load.
The numbers on the graphs have the same meaning as in Fig. 3. Except as indicated below the resistors are from normal factory production.
(iii)(*a*) and (iv)(*a*) unimpregnated.
(iii)(*b*) special vacuum impregnation.

Performance was satisfactory when the resistors operated continuously under substantial load, but rate of failure was sometimes high during storage unloaded. Especially rapid failure occurred under light d.c. load whilst the resistors were subjected to RCS/11 tropical exposure cycles, the d.c. load being insufficient to cause any appreciable heating and drying out.

(3.1) Moist Heat Cycling Treatments, with and without Light D.C. Load

Eight different makes of resistor were obtained, some in both standard and improved types. Most were of 6-watt rating, but

Table 1

OPEN-CIRCUIT FAILURES OF VITREOUS-ENAMELLED WIRE-WOUND RESISTORS UNDER MOIST HEAT CONDITIONS

Group	Make	Number tested	Resistance	Resistance of enamel in immersion test	Number of moist heat cycles to cause indicated proportion of failures: With light d.c. load, 30%	With light d.c. load, 50%	Unloaded, 30%	Unloaded, 50%
			kilohms	ohms				
1 Crazed enamel	F	52	8·1 to 100	1×10^5	17	33	310	390
	G	12	15	4×10^5	7	10	40	65
	H (after thermal shock)	10	27	3×10^5	2	7	155	205
	J	16	20	7×10^4	27	35	10	65
	K	16	15	5×10^4	17	25	20	40
2 Crazed enamel	E	36	10 to 39	1×10^5	350	410	>570	—
	K (new type)	20	15	1×10^5	340	>550	>550	—
	L	10	50	1×10^5	338	347	>575	—
3 Uncrazed	E (new enamel)	20	15	2×10^{11}	>680	—	>680	—
	F (special enamel)	40	8·2 and 20	5×10^9	462	>680	>680	—
	H	74	27 and 50	1×10^{11}	>680	—	>680	—
	I	54	20	1×10^{10}	>680	—	>680	—

some were 3 watt. Samples of all the resistors were subjected to the following cycling treatment:

(*a*) Sixteen hours in saturated air at 55° C.
(*b*) Eight hours in ordinary room atmosphere.

(*a*) and (*b*) comprised one cycle and the treatment was repeated until failure of the resistor occurred. Half the samples subjected to cycling treatment were unloaded and half were given light d.c. loading (0·2% of the nominal Inter-Services power rating) during period (*a*) only.

The number of cycles necessary for 30% and 50% open-circuit failures are shown in Table 1, where resistors of the same make and type are grouped together irrespective of value or rating. Where failures occurred they were generally more rapid with applied direct current than unloaded. In the Table the resistors have been divided into three groups. In group 1 they fail rapidly both with light d.c. load and unloaded (50% became open-circuted within 35 cycles when loaded and within 100 to 300 cycles when unloaded). Group 2 show a large proportion of failures (30–50%) after 300 to 400 cycles with light d.c. load, but show few failures up to 600 cycles unloaded. Group 3 show few failures up to 680 cycles even when loaded with light direct current. Severe thermal shock (see Section 3.6) transferred make H from group 3 to group 1.

A definite difference between resistors was that those in groups 1 and 2 were crazed or porous, as could be seen under suitable lighting conditions or as revealed by special techniques (see Section 3.2), whilst those in group 3 were uncrazed. Crazing could arise during manufacture from failure to match the coefficients of expansion of the glaze and the porcelain former.

These results suggest that absence of crazing or other discontinuities in the glaze would ensure long shelf life and satisfactory performance under light d.c. load in damp situations. It is possible, however, for some crazed vitreous enamels to give reasonably good performance; the difference between groups 1 and 2 probably depends upon the chemical nature of the glaze.

(3.2) Tests of Faulty Vitreous Enamel Coating

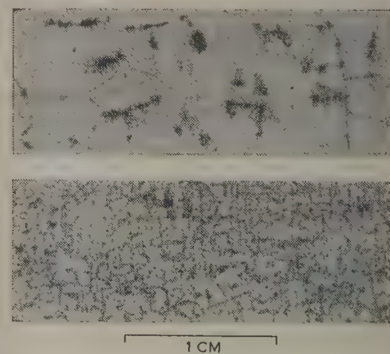
Cracks or other faults in vitreous coatings on resistors, although not always easily visible even under the microscope, are readily checked by measuring the insulation resistance of the coating in damp atmosphere.

In one method of test the resistance was measured between a metal-foil electrode pressed into contact with the coating and the resistor winding after conditioning in damp air.

In another method the resistor was supported vertically with the lower wire and terminal dipping into carbon tetrachloride. A layer of tap water of suitable depth was then poured on top of the carbon tetrachloride, so that the upper wire and terminal of the resistor were in the air and out of contact with the water. The depth of the water layer was the maximum compatible with isolation from the connecting wires and terminals. Resistance was measured between a metal plate immersed in the water layer and the upper terminal wire of the resistor.

The resistances obtained were always much lower when the vitreous glaze contained cracks or pinholes than when the coating was continuous, and were lower in the immersion test than when merely conditioned in damp air. In the immersion test, resistances in group 3 were all greater than 10^9 ohms whereas all those in groups 1 and 2 were less than 10^6 ohms (see Table 1, column 5).

The craze pattern on a resistor can be readily revealed by a procedure which makes use of the sensitivity of photographic emulsions to direct current. A piece of sensitive gelatine film is placed in contact with the surface of the glaze, a slightly smaller piece of tinfoil is placed over the gelatine film, and strip of rubber is held in tension against the tinfoil so as press the gelatine into firm contact with the glaze. After co ditioning to equilibrium in moisture-saturated air, a dire voltage is applied between the tinfoil and one of the termina of the resistor. Current then flows from the tinfoil, through t gelatine film and down the surfaces of the cracks to the resist winding. The current concentration in the cracks and oth faults causes a corresponding precipitation of metallic silver the photographic emulsion, which becomes more intense development. Fig. 7 shows enlarged photographs of the imag



(*a*)

(*b*)

Fig. 7.—Craze pattern on vitreous-enamelled wire-wound resistors
(*a*) Make F, 47 kΩ, 6 watts.
(*b*) Make G, 15 kΩ, 6 watts.

obtained with two resistors, one of make F (standard ename and the other of make G. The process gives an accurate ma of cracks and other faults in the glaze.

(3.3) Cause of Failure with Light D.C. Loading

When vitreous-enamelled wire-wound resistors are operate under light d.c. load in damp atmosphere, leakage currents flo between the terminals along the surface of the glaze. The cor struction of these resistors is, however, such that potenti differences can exist between the glaze surfaces and the wi immediately underneath. If cracks or pinholes exist, sma steady currents flow along the fault and may lead to electr chemical corrosion of the wire. This usually occurs near th anode end of the wire. Failure during moist storage withou applied load arises from the same cause, the responsible poter tials being contact e.m.f.'s.

(3.4) Operation under Full Load

Samples of makes E, F, H, I and K, all from standard pro duction, and of makes E and F with newly developed ename were tested under continuous full load (50 c/s) with the resistor supported horizontally in air at room temperature. A total o 48 resistors, all within the range 8·2–39 kilohms, were testec All withstood this loading for 4 150 h without failure.

There is evidence that behaviour under heavy d.c. load ma be less satisfactory, especially if the alkali content of the glaz or ceramic former is appreciable. Fig. 8 shows a large vitreous enamelled wire-wound resistor which has suffered serious corro sion leading to open-circuit after a relatively short period o operation under d.c. load. The corrosion products whicl formed at the cathode were highly alkaline. For satisfactor performance the sodium content of the former and glaze shoul be a small fraction of 0·1%.

Fig. 8.—Electrolytic corrosion near cathode terminal of vitreous-enamelled wire-wound resistor.

(3.5) Intermittent Loading

The above 48 resistors, which had been fully loaded for 4 150 h ithout any failure, were then subjected to full-load on-off cycling ive minutes on, five minutes off) under the same ambient con-itions without interruption except for measurement.

After 5 200 h of cycling (31 200 load cycles) all except four till survived. Resistors which were initially uncrazed (half those ested) had nearly all become crazed after 1 000 h of cycling and eveloped a low insulation resistance in the water-immersion est (Section 3.2).

(3.6) Thermal Shock

Sudden cooling from 230° C to 0° C caused crazing of the itreous enamel on resistors H and I. These were previously ncrazed and had successfully withstood several hundreds of noist-heat cycles with light d.c. load.

A number of resistors of make H were given the above thermal-hock treatment, and then subjected to moist-heat cycling, half vith light d.c. load and half unloaded. Although the shock reatment caused no immediate open-circuit, failure under moist-heat cycling was then relatively rapid, especially with light d.c. oad (see Table 1). This resistor is, in fact, transferred from roup 3 to group 1 by the shock treatment.

(3.7) Resistance Changes during Test

Periodically during test, resistances where open-circuit had not ccurred were measured after reconditioning to room atmo-phere. After the humidity cycling treatments resistance changes vere generally small, not significantly different for the unloaded nd lightly loaded samples, and with one exception showed no endency to drift in either direction. Table 2 shows mean hanges of value (ignoring the sign) after 84 humidity cycles. Aean values were all within 0·3% of the initial values except or make K (new type), which had increased by 1·8%.

Changes during full-load continuous and intermittent opera-ion were in some cases several times greater than during umidity cycling (see last two columns of Table 2).

Table 2

CHANGES OF RESISTANCE OF VITREOUS-ENAMELLED WIRE-WOUND RESISTORS AFTER DIFFERENT TREATMENTS

Make	After 84 humidity cycles		Mean change of resistance	
	Mean change of resistance	Proportion of resistors showing deviation exceeding 0·5%	After 4 000 h under full load	After 4 000 h under full load plus 5 000 h intermittent loading
	%	%	%	%
E	0·26	9	0·05	0·04
E (new) enamel)	0·06	5	0·08	0·00
F	0·3	9	0·4	0·7
F (special enamel)	0·2	15	0·4	0·5
G	0·02*	0	—	—
H	0·15	2	0·9	1·7
I	0·2	11	0·1	0·3
J	0·2*	0	—	—
K	0·3*	25	0·7	0·9
K (new type)	1·8	100	—	—
L	0·2	0	—	—

* Only 25% of the total number of samples tested in these groups survived 84 humidity cycles without open-circuit failure.

(4) HIGH-STABILITY CARBON-FILM RESISTORS (GRADE 1)

High-stability (grade 1) fixed resistors consist of a thin layer of carbon deposited on a porcelain former by a process of cracking a hydrocarbon vapour such as methane. A helical groove is subsequently cut in the surface of the former, the thickness of the carbon deposit and the width of the track determining the resistance. The delicate track is usually protected by a varnish coating but some are enclosed in an insulating tube.

As their name implies, such resistors are usually very stable: their resistance is not moisture-sensitive and they withstand electrical loading without large drift of value. Failure when it does occur is a fairly sudden open-circuit, but it is usually considered that if these resistors withstand 100 h full load without failure, life under load is very long. Braner and Easterday[10] found that with 237-kilohm deposited-carbon resistors (critical value*) operated at full rated load at 70° C, mean change of value after 12 000 h was between ±0·5 and ±1% but occasional individuals showed much greater changes. Borocarbon resistors (100 kilohms) were less stable.

Unlike the behaviour of grade 2 carbon-composition resistors, the failure rate of grade 1 resistors during full-load testing is more rapid with d.c. than with a.c. loading, and modern specifications call for d.c. loading in the endurance test. As with wire-wound resistors, the alkali content of the ceramic former must be low if electrolytic deterioration under load is to be avoided.[11]

Rapid failure during tropical exposure with light d.c. loading (0·2% of the rated voltage) has been reported from a number of Government Establishments. In some tests on values of 10 kilohms and above, all failed within seven days, with both insulated and uninsulated types. In other tests failure was confined to higher values (1 megohm and above) and occurred after periods of 14, 28 or 84 days. Failure took the form of open-circuit or large increase of value. In some other reported tests, behaviour under these conditions was satisfactory,

* 'Critical value' is the maximum resistance which will dissipate the nominal loading without exceeding the maximum permissible voltage across its terminals.

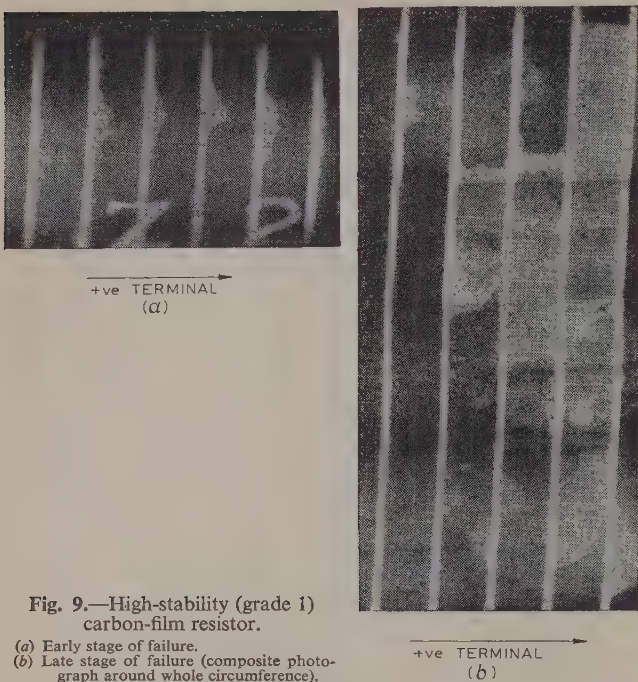


Fig. 9.—High-stability (grade 1) carbon-film resistor.
(a) Early stage of failure.
(b) Late stage of failure (composite photograph around whole circumference).

especially with insulated types. Under light a.c. loading very few failures occurred after 84 days' tropical exposure and no failures occurred with unloaded resistors.

Manufacturers claim that this test with light d.c. load is too severe and does not represent practical conditions of use, where performance is satisfactory.

The E.R.A. has investigated the cause of failure under these conditions, and studied the reasons for the marked differences in behaviour in different testing laboratories.

It was evident from the results of this investigation that failure under these conditions was electrochemical in nature, and that an anodic reaction could cause oxidation of the carbon film to carbon dioxide under suitable conditions and eventually open-circuit the track. Fig. 9(a) shows track erosion in its early stages in a 100-kilohm 1·5-watt resistor. Open-circuit and removal of considerable portions of the resistive track may follow [see Fig. 9(b), which is a composite photograph extending around the whole circumference of a resistor].

Conditions conducive to rapid carbon erosion of this nature exist where condensation of water droplets on the resistor can occur and where the protective film of paint or varnish contains faults such as pinholes, scratches or dust particles or the film is moisture-absorbent. Differences in performance which have been obtained in different testing establishments are due to varying degrees to which condensation occurs on the resistors. Testing chambers in which high humidity is obtained by steam injection, or by maintaining a bath of water at a higher temperature than that of the resistors under test, impose particularly severe conditions. Complete protection of the resistive element by enclosure in an insulating sleeve and proper sealing at the ends is advisable if it is desired to avoid this hazard.

Metal-film resistors in which a ceramic or glass base is coated with a thin film of a metal alloy, e.g. platinum-gold, may be subject to the same trouble unless moisture protection is complete. There is some evidence that metal-oxide-film resistors are more stable in this respect.

(5) CONCLUSION

The following general conclusions may be reached regarding failure mechanisms of the various types of resistors studied.

Carbon-composition resistors, both moulded and varnish-film types, fail under load by a slow thermal process. At the operating temperature the synthetic resin in which the carbon is dispersed gradually shrinks and improves the contact between the carbon particles, leading to continuous reduction in resistance whilst the load is maintained. At constant applied voltage this results in overheating and finally in catastrophic failure. However, many commercial resistors of this type remain within 20% of their initial value, even after two or three years' continuous operation under full load.

The rate of deterioration of cold-moulded types under load depends upon the completeness of the wax-impregnation process, on the type of wax used and on other details of the manufacturing process. Operation in an inert atmosphere which delays deterioration of the wax improves performance.

The resistance when unloaded or lightly loaded is sensitive to moisture and in some cases large changes may occur over long periods of time in humid air; these are largely reversible on subsequent drying. Lack of dimensional stability of the resistive element necessitates special care by the manufacturer in attaching the terminal wires.

Vitreous-enamelled wire-wound resistors operate satisfactorily under full-load conditions, provided that the alkali content of the former and glaze is low. The latter requirement is especially important under d.c. stress.

Failure may, however, be rapid during tropical exposure unloaded but especially when lightly loaded with direct current. This is due to electrochemical corrosion taking place at the base of faults in the vitreous coating. A crazed or otherwise faulty enamel should be avoided. Open-circuit failure during ordinary storage has also been reported and this probably arises from the same cause.

High-stability cracked-carbon resistors are not sensitive to changes in atmospheric humidity, i.e. resistance is not affected by humidity, but if they are operated under light d.c. load under very damp conditions they may suffer electrochemical attack and open-circuit. Reliability under these conditions depends upon the degree of moisture protection afforded by the varnish, paint or other covering applied to the resistive element. A complete seal is preferable. Condensation of water on this type of resistor should be avoided.

Other thin-film resistors may be similarly subject to electrochemical failure in wet atmospheres, and proper protection against moisture is necessary.

For satisfactory operation under full d.c. loading it is important that the alkali content of porcelain or glass formers used for thin-film or wire-wound resistors, and of the vitreous coating in the case of wire-wound resistors, should be low.

Short-term normal load tests on grade 2 resistors are of little value as an indication of ultimate life. Overload tests, if used for accelerating load-life, must be checked for the particular make and design of resistor under consideration. No general rules regarding the quantitative effects of overload are possible.

A light d.c. load test in damp atmosphere is regarded as a valuable type test for vitreous-enamelled wire-wound resistors which may fail even in storage as a result of electrochemical action. A similar test on grade 1 thin-film resistors is desirable where they are required to operate under these test conditions, especially under conditions of moisture condensation. However, they are not particularly prone to this type of electrochemical failure under milder operating conditions such as might cause failure of certain types of vitreous-coated wire-wound resistors.

(6) ACKNOWLEDGMENTS

The author wishes to thank the Director of the Electrical Research Association for permission to publish the paper, and the Ministry of Aviation for their support throughout the course of the work.

Acknowledgment is made to I. D. L. Ball, R. G. J. Butler, W. Plumtree, F. G. Rivers and J. J. Walsh who did much of the experimental work, and to C. G. Garton for valuable advice during the course of the work and the preparation of the paper.

The author is also indebted to many manufacturers who supplied samples for test and freely provided facilities to the investigators to study their manufacturing methods.

(7) REFERENCES

(1) Church, H. F., and Walsh, J. J.: 'Investigation of Deterioration of Moulded Carbon Resistors: First Report', E.R.A. Report Ref. Z/T92, 1953.

(2) Walsh, J. J., and Church, H. F.: 'Investigation of Deterioration of Moulded Carbon Resistors: Second Report', E.R.A. Report Ref. Z/T96, 1955.

(3) Church, H. F.: 'Grade 2 Fixed Resistors—Instability During Storage', E.R.A. Report Ref. Z/T113, 1958.

(4) Rivers, F. G.: 'The Performance of Insulated Carbon-Resin Film Resistors', E.R.A. Report Ref. Z/T95, 1955.

(5) Rivers, F. G.: 'The Performance of Insulated Carbon-Resin Film Resistors II', E.R.A. Report Ref. Z/T105, 1955.

(6) Church, H. F., and Walsh, J. J.: 'Study of the Failure of Vitreous Enamelled Wire Wound Resistors', E.R.A. Report Ref. Z/T107, 1956.

(7) Ball, I. D. L., and Butler, R. G. J.: 'High Stability Carbon-Film Fixed Resistors. D.C. Failure in Moist Atmospheres', E.R.A. Report Ref. Z/T124, 1959.

(8) Dummer, G. W. A.: 'Fixed Resistors' (Pitman, 1956).

(9) Davis, B. L.: 'Printed Circuit Techniques: An Adhesive Tape Resistor System', National Bureau of Standards Circular No. 530, 1952.

(10) Braner, H., and Easterday, J. L.: 'Life Characteristics of Carbon-Film Resistors after 12 000 Hours of Operation', *Transactions of the American I.E.E.*, 1959, **78**, Part I, p. 201.

(11) Burkett, R. H. W.: 'The Performance of Pyrolytic Carbon Resistors', *British Communications and Electronics*, **6**, p. 264, 1959.

DISCUSSION ON
'DESIGN OF AN AUTOMATIC SENSITIVITY CONTROL FOR A NEW SUBSCRIBER'S TELEPHONE SET'*

NORTH-EASTERN MEASUREMENT AND ELECTRONICS GROUP, AT NEWCASTLE UPON TYNE, 2ND NOVEMBER, 1959

Mr. J. A. Stanfield: To ensure satisfactory operation of a subscriber's telephone there are two limiting factors to take into consideration: one of transmission and the other of signalling. The transmission limit is stated in terms of local end-line resistance, as is the signalling limit. Presumably, therefore, the new telephone can be used up to the transmission limit of 1 kilohm only when the signalling limit allows this. Are there many exchanges where the signalling limit is anywhere near 1 kilohm?

There is a difference between the manufacturer's curves and the authors' curves in Fig. 3, the manufacturer's curves being more optimistic. Nevertheless it would seem that a worth-while improvement can be expected by changing to the new telephone on all lines having a resistance of down to 300 ohms.

On Post Office subscribers' telephones the side-tone balance network impedance must necessarily strike a mean of the various line impedances, but on direct point-to-point telephones such as exist in noisy situations in power stations the line impedance is constant. Therefore, now that the balance network is separate, it should presumably be possible to reduce side tone considerabl by adjustment of the network components.

I understand that a dial-locking device is available for use o this new type of telephone, and although not mentioned i the paper it would seem to be an invaluable guard agains unauthorized calls, especially in view of the coming subscriber trunk-dialling service.

Messrs. F. E. Williams and **F. A. Wilson** (*in reply*): In anticipa tion of the new telephone, all new automatic exchanges ordere after June, 1955, provide for signalling over 1-kilohm line Most of the older exchanges can be modified to meet the ne limit, and, in fact, a substantial number have already been s modified.

If our curves differ from those given by the manufacture it is probably due to different techniques of assessment. Th Post Office figures are based on subjective measurements mad with speech-testing crews.

It would be possible to modify the balance network to matc a particular line impedance on a direct point-to-point telephon circuit, and a substantial reduction in side tone could be achieve by so doing.

* WILLIAMS, F. E., and WILSON, F. A.: Paper No. 2867 E, July, 1959 (see 106 B, p. 361).

DISCUSSION ON
'THE RECOGNITION OF MOVING VEHICLES BY ELECTRONIC MEANS'*

NORTHERN IRELAND CENTRE AT BELFAST, 8TH MARCH, 1960

Mr. R. W. Adams: The problem of increasing traffic congestion is of as much concern to the passenger-transport operator in provincial towns and cities as it is in London. Since the bus electronic scanning indicator was announced in 1957, municipal operators have awaited with the greatest interest the results of these experiments. Only an organization operating passenger services on so large a scale as London Transport is in a position to undertake the whole development of such a project.

The ingenuity of the equipment devised to detect the passing of each scanning point by particular buses has been made very clear by the authors, but a number of questions suggest themselves to the transport operator who is principally concerned with what can be achieved by the use of the equipment. For example, how many routes would be taken care of by one controller? What number of other persons would the controller need to assist him directly, and what would be their duties in the cor trolling office, out on the routes and in other locations? In th connection it is presumed that there would be more than on controller for the 500 routes referred to in Section 2.

Does the introduction of this indicator allow of any reductio in the numbers of staff at present required for the operation c bus running control by the existing conventional methods?

Is the fact that only 200 scanning points are required fc 500 routes explained by the fact that very many of the route overlap? As a bus proceeds from one scanning point to anothe its progress can be checked. How many scanning points woul there be in each direction along a route?

How is any conclusion arrived at by a controller translate into action to remedy irregular running on the routes?

Has experience with the use of the indicator so far shown to give any improvement in the actual control of bus runnin which can be measured?

* PICK, T. S., and READMAN, A.: Paper No. 2758 M, October, 1958 (see 106 B, p. 186).

Is the scanning equipment used only at peak hours or will it : operated continuously throughout the operating hours? ow is power switched on, when required, at the scanning ›ints? Is it by remote control with the operating switch in the ›ntroller's room?

It is indicated in Section 3.5 that the voltage and currents in e transmitting circuits are within the limits acceptable for ›st Office lines. Will the Post Office, in fact, rent lines for ›eration of a system of this kind?

Can the authors give any idea of the cost of providing an stallation of this sort?

Messrs. T. S. Pick and **A. Readman** (*in reply*): It is early yet › say what the final form of control will be, but a single con-oller will handle a number of routes and will control by :lephone communication to strategic points along the route ›here staff are already stationed. The total number of staff involved is not expected to be any greater than is used to-day—in fact, is hoped to be less for a much greater control.

Where routes overlap, it is anticipated that scanning units will be able to deal with a number of routes, and the siting of the scanners will depend largely on how the information can be handled usefully by the controller. It is anticipated that no bus will indicate itself at intervals longer than about 10 min. Experience already gained on the 74 route has shown the value of the indication, the controller being aware of congestion or of a faulty vehicle before any verbal message can be obtained from road-side staff. This continuous supervision has resulted in an improvement in regularity.

It is expected that the Post Office will, in fact, rent lines for operation of a system of this kind. The cost of the system is small when related to the present cost of personnel for supervision.

PAPERS AND MONOGRAPHS PUBLISHED INDIVIDUALLY

Summaries are given below of papers and monographs which ave been published individually. The papers are free of charge; he price of the monographs is 2s. each (post free). Applications, uoting the serial numbers as well as the authors' names, and ccompanied by a remittance where appropriate, should be ddressed to the Secretary. For convenience, books of five ouchers, price 10s., can be supplied.

:apid Methods for ascertaining whether the Activity of a Weak Radioactive Sample exceeds a Predetermined Level. Paper No. 3258 M.

E. H. Cooke-Yarborough, M.A., and R. C. M. Barnes, B.Sc.(Eng.).

The paper considers methods of determining whether the mean ate of occurrence of random events, such as radioactive disintegra-ions, is greater or less than some given tolerance rate. The most fficient method is one which makes this determination in the shortest ime and with an acceptably low probability of error. Sequential test ›rocedures minimize the sample size by observing the random process intil some terminating condition is satisfied. The performance of everal sequential tests is investigated, with particular reference to :onditions encountered in monitoring radioactive contamination on he hands. The most efficient of these tests is shown to be one which neasures the difference between the number of events actually observed ind the number expected at the tolerance rate. Under conditions likely o be encountered in routine monitoring for radioactivity on the hands his method gives a fivefold saving in time over the method now in use.

\ Universal Non-Linear Filter, Predictor and Simulator which optimizes itself by a Learning Process. Paper No. 3270 M.

Prof. D. Gabor, F.R.S., W. P. L. Wilby, Ph.D., and R. Woodcock, Ph.D.

A machine is described which consists of a universal non-linear ilter, which is a highly adaptable analogue computer, together with a :raining device. The analogue machine has 18 input quantities from which it can compute in about 2·5 millisec 94 terms of a polynomial, each term containing products and powers of the input quantities, with adjustable coefficients, and can form their sum. The input quantities may be, for instance, 18 past samples of the values of a stochastic variable which is fed into the machine, and the result of the computation is an output function which contains 94 free variables. The training device optimizes the output by successive adjustment of the variable coefficients until it has approached a target function as closely as can be achieved with a polynomial of 94 terms, by the criterion of the least mean-square error. This is done by repeatedly feeding into the machine a record of the stochastic process, long enough to be representative, and adjusting the variable coefficients, one at a time after each run, by a strategy which ensures that the error will monotonically decrease from run to run.

In order to make the machine an optimum filter it is trained on a record of a noisy process, together with a target record which contains the signal only. It is taught as a predictor by taking as the target function a value of the stochastic process advanced by a certain time interval beyond the last value which goes into the input. It is trained as a simulator, for instance of an unknown mechanism, by feeding it with the input of the mechanism to be simulated at one end and presenting it at the other with its output as target function. The machine will then make itself into a model of the device to be simulated and the non-linear transfer function of the device can be read off from the final setting of the coefficients, as nearly as it can be represented by a 94-term polynomial. The machine is not confined to single-input systems.

The machine incorporates 80 analogue multipliers of a novel 'piezomagnetic' type which, in its present form, can perform over 1 000 multiplications per second with an error of 0·5% or less.

A few examples of the first test applications of the machine are described.

Extension of the Dual-Input Describing-Function Technique to Systems containing Reactive Non-Linearity. Monograph No. 383 M.

R. M. Huey, B.Sc., B.E., O. Pawloff, Dipl.Ing., and T. Glucharoff, Dipl.Ing., M.E.

Kochenburger's describing function and also the dual-input describing function due to West, Douce and Livesley are well known. These techniques enable the graphical solution of non-linear differential equations in which the non-linear coefficient is associated with the first-order differential term to be obtained. The paper describes an extension of these methods, allowing the non-linear coefficient to be associated with any term in the differential equation.

It is shown that the typical non-linearity presented by an iron-cored inductor or by a ferroelectric capacitor may be resolved into a simple feedback system containing a single non-linear element which is independent of frequency, together with elements possessing linear transfer functions. The resulting equivalent block diagram is suitable for analysis by either of the describing-function techniques. As an example, the dual-input technique is used to predict accurately jump phenomena in an iron-cored ferroresonant circuit, and the validity of the equivalent block diagram suggested is also checked by simulation on an analogue computer.

Intermodulation on Amplitude-Modulated Multi-Channel Line Links. Monograph No. 384 E.

J. C. H. Davis, M.A., and H. O. Friedheim, B.Sc.

The paper develops a method for estimating the total intermodulation noise power falling into any channel in the transmission band of an a.m. multi-channel line system with many repeater sections in tandem. These sections may be composed of widely varying lengths of cable imperfectly equalized by non-identical amplifiers whose in-band feedback, output load impedance and output network response all vary with frequency. The whole system is loaded in such a way that the signal levels at each intermodulation source are also functions of frequency.

After developing the basic theory, methods are given for estimating the system performance before building an amplifier and also when a model amplifier is available. These methods are more general than those previously published and are illustrated by examples.

The Resonance Excitation of a Corrugated-Cylinder Antenna. Monograph No. 386 E.

J. R. Wait, M.A.Sc., Ph.D., and A. M. Conda, B.A.

Radiation from an axial magnetic line or slot source on the surface of a corrugated cylinder is considered. It is indicated that the power radiated in a given mode for the structure depends critically on the surface reactance and the circumference of the cylinder. In fact for certain values of these parameters particular modes are strongly excited and contain most of the radiated power. Numerical results are presented for several interesting cases.

The analysis is extended to an elliptic cylinder whose surface also possesses an inductive reactance. In order to facilitate the solution it is necessary to assume a special azimuthal variation of the surface reactance. For the model as chosen, strong resonance characteristics are again obtained. This model may be adapted to study the problem of a corrugated panel on a flat metallic ground plane which is excited by a parallel slot source.

Signal Flow-Graph Analysis and Feedback Theory. Monograph No. 388 E.

R. F. Hoskins, M.Sc.

The solution of a system of simultaneous linear equations may be obtained by inspection of an associated system of nodes and connecting branches called a 'signal flow-graph'. This provides an alternative to conventional algebraic methods which is of particular interest in the case of network analysis since the flow-graph can be set up directly by inspection of the network without having to formulate the associated equations. In the paper the formal theory of flow-graph analysis is developed and applied to certain aspects of feedba[ck] theory, and it is shown that the classical results of Bode can be obtain[ed] and generalized relatively simply by this approach.

The Received-Amplitude Distribution produced by Radio Sources [of] Random Occurrence and Phase. Monograph No. 389 E.

W. C. Bain, M.A., B.Sc., Ph.D.

A theoretical calculation is given of the amplitude probabil[ity] distribution to be expected on an ionospheric v.h.f. forward-scat[ter] circuit due to reflections from meteor trails alone. The analysis [is] based on the addition of a large number of signals with a frequen[cy] of occurrence inversely proportional to the square of their amplitu[de] and will therefore apply to other problems in which this relation hol[ds]. The calculated distribution is compared with a small number [of] practical results, and a method is outlined for deriving the relati[ve] proportion of meteoric and turbulent-scattering components in t[he] signal.

A New Type of Piezo-Electric Flexural Vibrator in the Form of Balanc[ed] Cantilevers. Monograph No. 391 E.

Mrs. S. Ayers.

The flexural vibrator, designed to vibrate at about 1 kc/s, basica[lly] consists of identical cantilever arms extending from a common ar[m] to form a symmetrical element. Two distinct shapes have be[en] considered—the H and the 'zigzag'. Some of the H elements ha[ve] uniform cross-section while others are arranged to have most of th[e] mass at the free ends in order to reduce the frequency for a specim[en] of given length. The 'zigzags' have folded arms of any number [of] sections (increasing the number of sections reduces the frequency).

The theory of the various forms and their frequency equations a[re] derived. Conditions for perfect balance of the reactions at t[he] supports are discussed.

Measurements have been made on H and 'zigzag' forms made fr[om] quartz slices ZYbΦ($\Phi = 0 - 10°$) and on 'zigzag' form fr[om] EDTXY$ltl\Phi$, 90°, 90°. Frequency/temperature behaviour, Q-fact[or] and displacement patterns of the elements are compared with theo[ry]. Since some of the conventional driving methods proved unsatisfacto[ry] a short Section is included on circuits.

Chain Codes and their Electronic Applications. Monograph No. 392 [E.]

F. G. Heath, B.Sc., and M. W. Gribble, B.Sc.

A type of binary digital code is described which is easily generat[ed] by computer circuits. The important properties of these codes a[re] described, and various electronic applications enumerated.

PROCEEDINGS OF THE INSTITUTION OF ELECTRICAL ENGINEERS

Part B. ELECTRONIC AND COMMUNICATION ENGINEERING (INCLUDING RADIO ENGINEERING), JULY 1960

CONTENTS

Bibliographical References.—It is requested that bibliographical reference to an Institution paper should always include the serial number of the paper and the month and year of publication, which will be found at the top right-hand corner of the first page of the paper. This information should precede the reference to the Volume and Part. *Example.*—SMITH, J.: 'Reflections from the Ionosphere', *Proceedings I.E.E.*, Paper No. 4001 R, December, 1954 (102 B, p. 1234).

Published by The Institution, Savoy Place, London, W.C.2. *Telephone*: COVent Garden 1871. *Telegrams*: 'Voltampere, Phone, London.'
Printed by Unwin Brothers Limited, Woking and London.